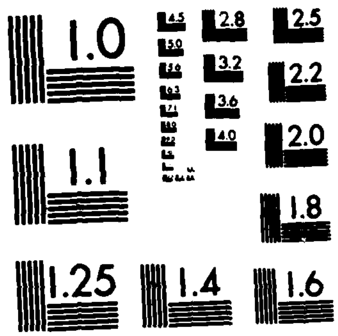
AD-A126 015 SOCIOECONOMIC IMPACT ASSESSMENT: COMMUNICATIONS 1/7
INNUSTRY PHASE III TECHNO..(U) ACUMENICS RESEARCH AND
TECHNOLOGY INC BETHESDA MD 02 FEB 79 FAA-APO-81-11-4
UNCLASSIFIED DOT-FA78WAI-932 F/G 17/2.1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

A126015

1

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FAA-APO-81-11 (4 of 6)	AD-A126 015	

4. Title and Subtitle	5. Report Date
Socioeconomic Impact Assessment: Communications Industry Phase III--Technology Forecast	February 2, 1979
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
9. Performing Organization Name and Address Acumenics Research and Technology, Inc. 4340 East-West Highway Suite 808 Bethesda, Maryland 20814	10. Work Unit No. (TRAIS) 11. Contract or Grant No. DOTFA78WAI-932
12. Sponsoring Agency Name and Address Department of Transportation Federal Aviation Administration Office of Aviation Policy and Plans Washington, D. C. 20591	13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code APO-210

15. Supplementary Notes

16. Abstract

DTIC FILE COPY

DTIC ELECTE MAR 22 1983 S D B

83 03 04 085

17. Key Words	18. Distribution Statement
	Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
		618 ~~610~~	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

The contents of this report reflect neither a position nor an official policy of the Department of Transportation. This document is disseminated in the interest of information exchange. The United States Government assumes no liability for the contents of this document or use thereof.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

Chapter One

Introduction

1. Introduction to the Unconstrained Technology Forecasts

1.1 Aviation Communications as a Segment of Telecommunications

In previous phases of this study, the historical evolution of aviation communications in the National Aerospace System (NAS) of the United States was reviewed and its present composition was described in detail. As a basis for forecasting the future development of this communications network, a set of alternative future scenarios for the national social and economic climate were defined, and for each scenario the size and composition of the United States aircraft fleet was forecast as a function of time. From these statistics, the future demand for aviation communications can be derived. In attempting to meet these needs, aviation communications will be founded by the available communications technology and constrained by the many factors governing the diffusion of new technology into the NAS commmunications network. Many of these diffusion factors have been discussed. They are strongly influenced by policy and planning decisions within the FAA, for whom this study is intended. The level of technology available for use in aviation, on the other hand, will be relatively independent of FAA decisions. This section seeks to provide a basis for forecasting the development of this broader communications technology.

Aviation communications is a highly important but nonetheless a relatively small part of the national communications system. As such, most of telecommunications technology will develop independent of aviation needs. Significant new technological developments are usually the result of basic research on materials and processes which are subsequently utilized in practical devices and then finally deployed in working systems. The emerging fiberoptic communications systems are an example of this materials-devices-systems development path. In areas having broad application to meeting telecommunications needs, such as the development of improved switching machines, sufficient funding will likely be available in the private sector to support the necessary work in all phases of this development path. In such areas, aviation will play the role of a user. In other narrower areas, government funding will play a more important role, especially where military needs are involved. Military avionics can be expected to see early application of may new developments in basic technology and will play an important role in specific areas. These military systems will be very advanced, high in cost, and not immediately applicable to civilian avionics, although some of their concepts and much of their technology will eventually be applied in the civilian sector. Some new technology will likely see application first in such military systems, followed by its adoption in lower-cost forms by the higher-volume communications areas and, in the aviation sector, by the commercial air carriers. Subsequently, its application in

general aviation will occur. Thus, while both private and government funding will be important in the overall telecommunications industry, general aviation will not be a major driving force behind its development, nor will the aviation communications market as a whole. This situation is depicted in Figure 1.1.

The forecasts of telecommunications technology will thus be largely independent of aviation and will be dependent on much broader factors. Once developed, aviation will provide an important market for the technology and will act as a user. The application of this technology to specific aviation systems will be described in later sections of this report as will the possible impacts on the national airspace system.

1.2 The Organization of Telecommunication Systems

Before proceeding further, it is important to define the elements of any commnication system. In this discussion we will use the description adopted by Shannon (1) and subsequently used elsewhere (2), in which a communication system is viewed as consisting of five parts as shown in Figure 1.2. First, there must be a source, which originates the message to be communicated. The source might be a person or could be a machine operating under the control of a stored program. Second, there must be a transmitter, which converts the message into a form suitable for transmission. The transmitter might, for example, convert speech energy into an analog-or-digitally-encoded electrical signal. Third, there must

Aviation Communications as a Segment of Telecommunications
(Arrows Indicate Technology Flow)

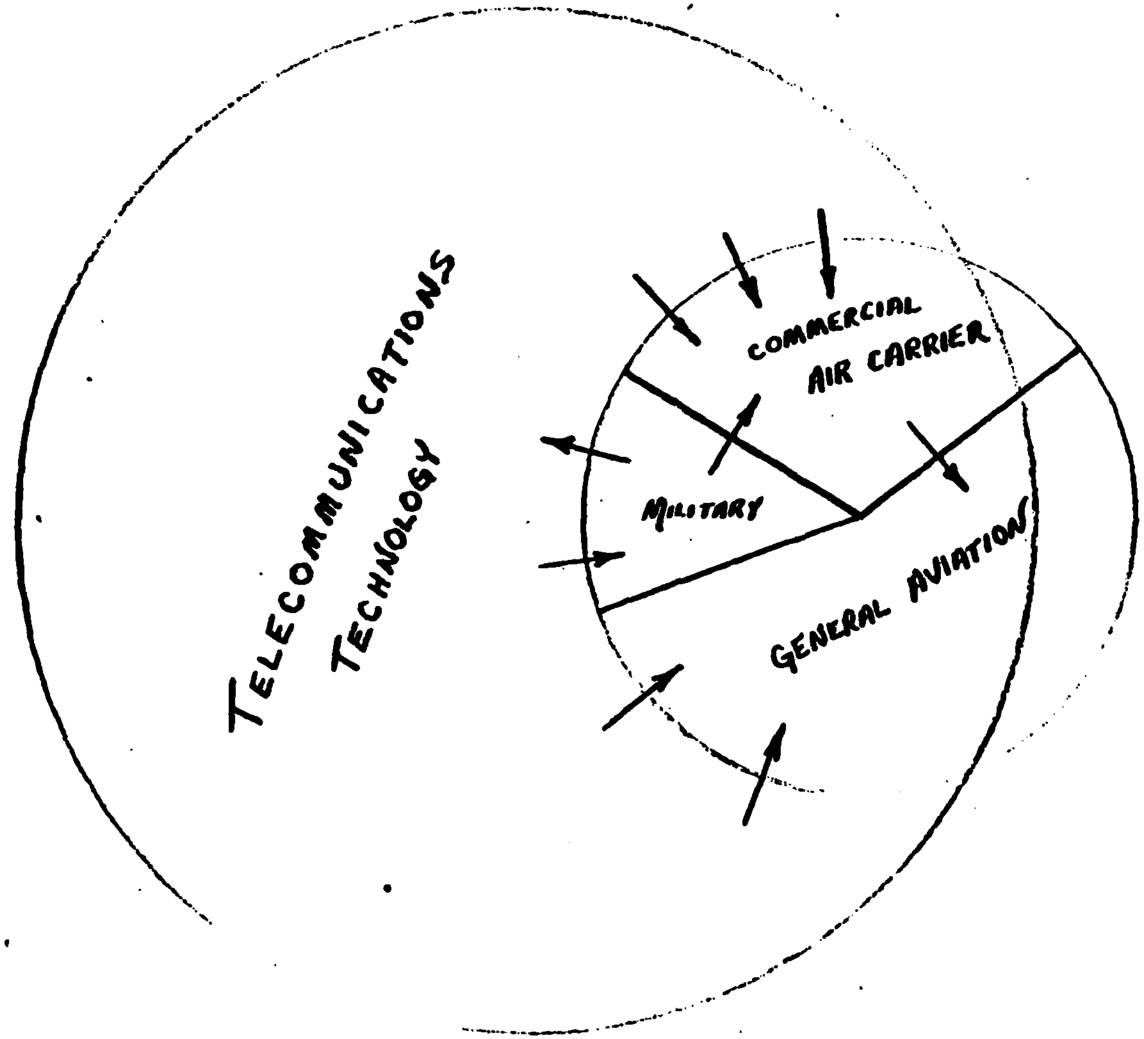


Figure 1.1

Elements o General mmunicat System

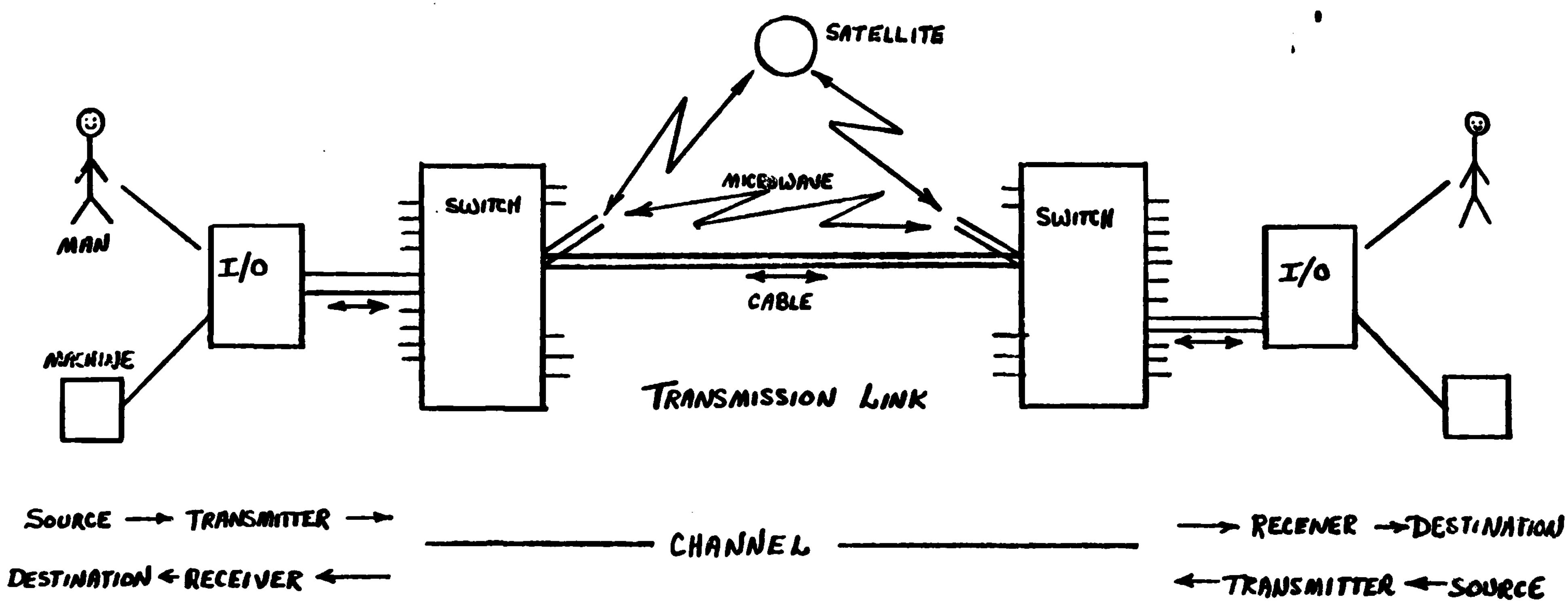


Figure 1.2

be a channel through which the message is transmitted. This might be a wire, a microwave link, or an optical fiber, for example. Fourth, a receiver is required to accept the message from the channel and convert it into a form which is comprehensive to the fifth and final part, the destination. The definition is the person or machine to which the message is to be communicated. While the exact boundaries between these different parts of the system are somewhat arbitrary, any communication system can be divided into such parts. This system description is consistent with the broad definition of communications as "the ability to transfer energy or information from one point to another" (3).

Most of the communications systems we will deal with will be two-way, allowing communications in either direction, and in such systems the source may also serve as a destination. This is certainly true of most ground-based systems. A possible exception is the national weather information network, which is primarily one-way, distributing weather information gathered by a distributed network or sensors. Even this network becomes two-way if inputs from airborne observers and the ability to be interrogated are considered. Some collision avoidance systems can also be considered two-way, since in one direction the radar signal is sent out while in the other, the plane sends back a return signal allowing its presence and position to be determined. In general communications, the telegraph and telephone are familiar examples of two-way systems, while television is one-way.

In order to forecast developments in future communication systems, it is convenient to divide the system somewhat differently from the five-part organization described above since the source and destination roles are usually found in one entity, which may or may not include some or all of the receiver and transmitter functions as well. For our purposes, we will divide general communication systems into three broad areas: input-output devices, switching devices, and transmission. Each of these areas will now be discussed briefly.

1.2.1 Input-Output Devices

Input-output devices perform the functions of the source and destination in the overall communication system. They provide a typically two-way interface between the system users. Teletypewriters and video terminals are common examples of input-output devices having keyboard or paper tape entry for information and providing a visual alphanumeric hard-copy or soft-copy display for the user. Note that these devices are the final source and destination machines in the communications path, although the actual source and destination devices may be the human users. Notice also that these devices convert the information into electrical signals suitably encoded for transmission and display of such information in printed form for the user, thus accomplishing at least a portion of the transmitter and receiver functions as well. The telephone accepts sound information and encodes it electrically for transmission as well as converting electrical signals into

sound information and therefore falls in a similar category. In the past, communications were nearly always between people in such systems, however, to an increasing extent, interaction between people and machines or between two machines will define the communications system. In many such systems, the input-output device may become increasingly difficult to recognize but will remain an important part of the overall system. The human interface will adopt increasingly sophisticated visual displays as well as advanced audio and perhaps tactile interfaces in both directions. The development of voice input, as well as voice output, capabilities will be an important part of these man-machine communication systems.

Perhaps more than any other area, the area of input-output devices is expected to see rapid growth. This segment of the communications industry is already characterized by a wide range of companies, from small entrepreneurial organizations capable of rapidly applying the most recent advances in technology to large, highly-structured telecommunications companies, involved not only in the development of new terminals but in the basic research necessary to solve the visual and auditory interface problems anticipated for future machines. The approaches used in the input-output area will likely span as wide a range as will the technologies employed. Developments in the visual display area will likely involve work on thin, color cathoderay rubes (CRTs) as well as flat panel displays using new display approaches. Auditory

man-machine interaction will be enhanced by the development of algorithms capable of recognizing human speech and accurately interpreting a substantial vocabulary. Telephones and other voice-input devices will become increasingly digital, employing sophisticated CODECs (coders-decoders) to perform analog-to-digital conversion in the input-output device itself. Microcomputers will provide such terminals with a substantial amount of stored-program intelligence and will work in concert with high density bulk memory devices capable of storing information on magnetic disks or monolithic bubble memories. Hard copy will be provided on many systems, making these terminals applicable to the huge market areas now served by newspaper, mail, and library functions. The challenge in the input-output area is to achieve high performance and versatility at a cost low enough to make them affordable in every home. The television sets of the year 2020 may in reality be sophisticated input-output devices complete with all of the features mentioned above. Present activity in video recorders for the home market is only the first step in expanding the features of traditional television receivers.

The input-output area is expected to be extremely active throughout the next quarter century. Competition among manufacturers is expected to be keen as this facet of the telecommunications industry races to solve the man-machine interface problem and allow developments in transmission and switching to be utilized by an increasingly information-oriented society.

1.2.2 Switching Devices

The first automatic switching devices which began to replace operator-switched telephone traffic in the early 1900's were electro-mechanical "step-by-step" devices controlled by the dial pulses from the input-output device itself (4). These switching devices grew in size and sophistication as automatic dialing of both local and toll calls spread across the nation. With the development of digital solid-state electronics, the concept of using a dedicated special purpose computer to control call switching and routing was born and the first such system was tested by the Bell System in 1960 (5). Electronic switching expanded rapidly throughout the 1960s and 1970s. These machines consist of a line interface to the transmission network, an electronic concentrator, the switching matrix itself (consisting of several stages) and the computer control section as shown in Figure 1.3. The technology of the stored program control section is that of any large computing machine, complicated in both hardware and software by the high requirements on reliability and the need for sophisticated built-in diagnostic, error detection, and maintenance routines. Line interface problems, complicated by the nature of the real-world transmission network (including lightning strikes, enormous in-band power line interference, and the necessity of remaining compatible with electro-mechanical input-output devices), have forced the line interface to remain relatively expensive and bulky on a per-line basis and until the late 1970s forced the switching network

Elements of an Electronic Switching System

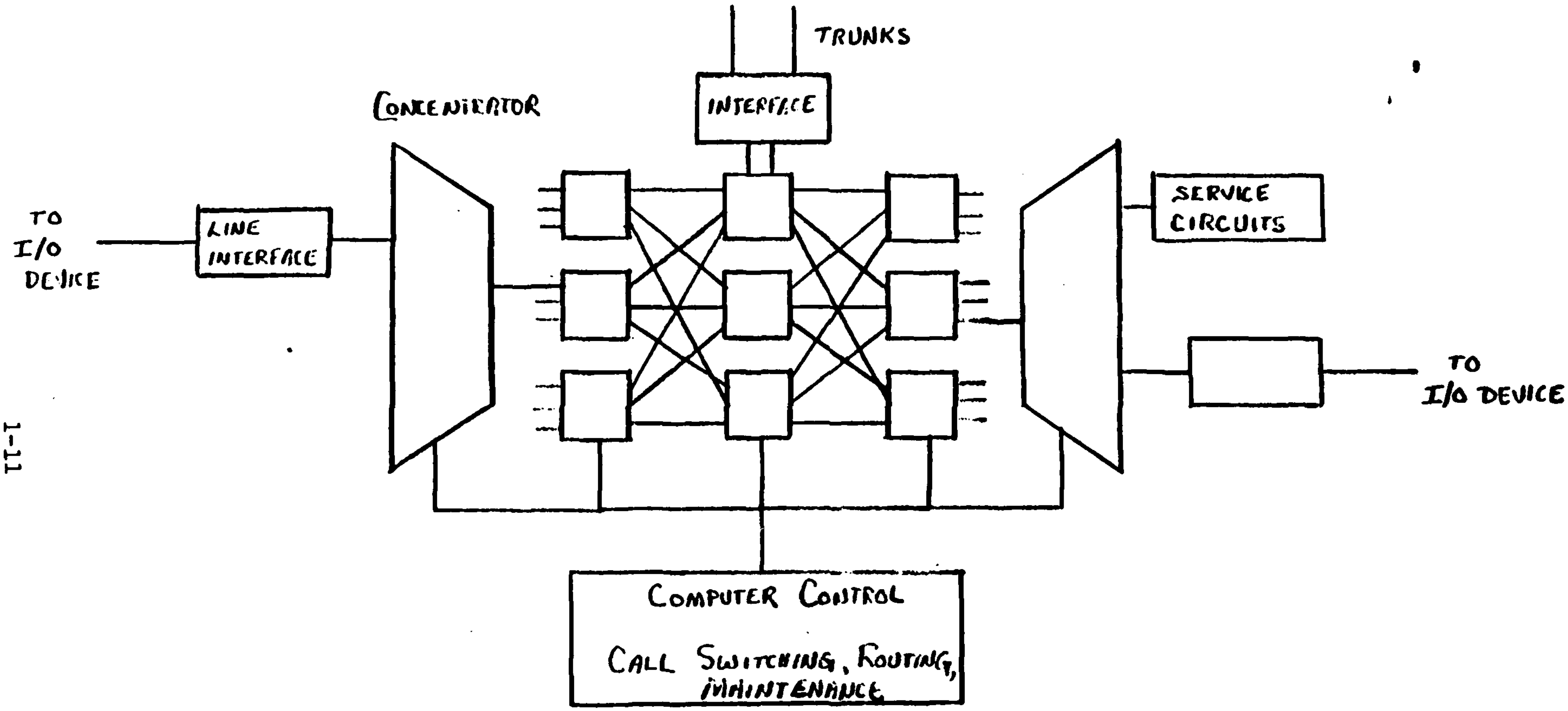


Figure 1.3

itself to remain electromechanical (relays). The development of inexpensive and reliable solid-state cross-points during the 1970s were important to the further growth of these "space division" electronic switches.

The telephone switching network has evolved into a complicated hierarchy of offices, ranging from large toll switching centers to small local offices interfacing with subscriber lines. Early switches were virtually all space-division, handling analog information on a dedicated path set up through the switch. As digital electronics expanded, time-division switching developed, in which many messages are handled via digital multiplexing in time on a single path, each message being assigned a specific time slot. Cross-points in a space-division system are replaced by logic gates in a time-division system, making the network itself more compatible with large-scale integration along with the control.

The switching area involves the development of highly complex high-speed computer-controlled networks capable of handling large volumes of information with very high reliability. Hardware -- consisting in part of highly-integrated logic and memory -- is only part of the challenge. Software routines capable of high-speed execution for call handling and maintenance are perhaps a more difficult developmental task along with the definition of machine architectures capable of high capacity and high reliability. The switching area promises to remain of critical importance as the

demands of society for increasing amounts of information grow and the capabilities of both the input-output and transmission areas expand.

1.2.3 Transmission Systems in Telecommunications

While the input-output devices encode the information into an electrical signal suitable for transmission and the switching systems route these signals to the correct destination, it is the responsibility of the transmission system to pass the information from source to destination without distortion. The earliest transmission systems were a pair of wires connecting the source and destination by way of one or more operator-controlled switchboards. Mounted on telephone poles, these wires had a diameter of about one-eighth of an inch and attenuated the signal amplitude about 0.1 db per mile (2) (a factor of two every 30 miles or so). It quickly became impossible to use a physically separate pair of wires this large for every transmission path, so the wire pairs were insulated and placed in a cable. Wire diameters decreased to about 0.02 inches resulting in many conductors per cable but with increased attenuation (approximately a factor of two every mile). To periodically amplify the signals to their original levels, repeaters were placed at approximately four-mile intervals. Since these repeaters are typically one-way devices, separate wire pairs were used for each direction, resulting in four-wire transmission paths. As the demand for more communications grew, a means was sought to carry more than one message per

wire pair. For voice signals (which predominated), a bandwidth from 200 to 3500 Hz ("voice band") is required. A single wire pair has a much broader bandwidth, however, and by translating the voice-band signals to higher frequencies for transmission, frequency-dependent, it became necessary for the repeater amplifiers to not only amplify reliably at high frequencies, but also to equalize the cable loss by having a gain-frequency characteristic which matches the line loss. Coaxial cables reduce crosstalk and loss, extending the useful frequency range and cable capacity into the megahertz range for long distance transmissions. The L3 coaxial system operating with a repeater spacing of four miles and a frequency band from 0.308 to 8.34 MHZ is capable of handling 1860 telephone channels or 600 telephone channels plus a 4.2 MHZ broadcast television channel on a pair of 0.375 inch-diameter coaxial tubes (2).

The transmission environment is noisy, resulting in both a theory of communication in the presence of noise (1) and in the search for modulation techniques capable of utilizing the transmission medium efficiently and transmitting the information reliably in the presence of noise. Continuous-wave approaches evolved amplitude, frequency, and phase modulation techniques and many sub-variations. Pulse-modulation techniques were explored and found better suited to some media and as they developed, time-division multiplexing was born. In this mode, pulse sequences representing different input signals can be interleaved in time

and transmitted over the same system. Pulse-amplitude, pulse-code, and delta modulation are all examples. As input-output devices and switching systems become more complex, time-division-multiplexing (TDM) is being used more and more.

In an effort to develop methods for signal transmission having lower cost, alternatives to cable have been pursued. After World War II, microwave radio was the most rapidly expanding transmission medium, partly due to the rapidly expanding television industry. Several frequency bands between 3 GHz and 12 GHz were selected. Since microwave transmission is a line of sight technique and towers are limited to a height of a few hundred feet, repeater spacings of the order of 25 to 30 miles are dictated by the curvature of the earth. Antennas and repeaters were developed capable of handling large traffic volumes. The Bell System TH system provides six channels in each direction, each channel carrying 1860 telephone circuits (2).

With the development of geostationary communications satellites (satellites which rotate with the earth, remaining fixed above a specific point on the ground), it has become possible to circumvent normal requirements for a 25-30 mile repeator spacing. In fact, a single satellite antenna can cover all 48 States plus Alaska and Puerto Rico. Such satellites contain an antenna, amplifiers, multiplexers and demultiplexers consistent with their role as repeating stations. The first domestic commercial U.S.

satellite, Westar I (Western Union) was launched in April of 1974 (6) and carried 12 amplifiers, each with a 36 MHZ bandwidth capable of carrying 1200 voice channels, one color TV signal, or data at 50 M bits/second. Like most U.S. satellites, a 500 MHz bandwidth centered around 6 GHz is used for the up-like (satellite receiving) and a similar bandwidth centered around 4 GHz is used for the down-link (satellite transmitting). Higher frequencies and wider band-widths are being developed for future satellites using carriers at 12/14 GHz and higher. These would avoid some interference problems with terrestrial microwave links and allow greater traffic capacity. Satellite orbit separations, positioning ground stations, and a variety of other concerns are part of the satellite communications picture, which is driven as much by non-technical considerations as by technical capabilities.

Microwave radio has a limited frequency spectrum available and in an effort to achieve greater traffic capacity, work is underway to develop techniques which use the existing spectrum more efficiently as well as to develop alternative transmission media. Two media which hold promise for the future are millimeter waveguides and fiber optics. Waveguides approximately two-inches across operating in the 50-100 GHz range have the potential of carrying many times the traffic of a microwave link (of the order of 100,000 telephone channels); however, this approach is not without its problems. New devices are required to work at these higher frequencies. The waveguides must also be precisely aligned

to avoid undesired propagation modes. Such high capacity has another problem as well -- in order to use the capacity it may be necessary to collect the traffic from widely distributed sources. In some cases, the job of collecting such high concentrations of traffic may be more formidible and expensive than the transmission system itself.

Probably the most active area of transmission at present is the area of optical fibers (7-8). In fiber-optic communications, the electrical signal (suitably encoded) is first converted into an optical signal having a wavelength in the visible to near infrared range (0.8-1.1 um). This signal of varying light intensity is carried over small glass or plastic fibers having diameters of a few microns. The fibers are clad with a covering having a lower refractive index so that light is contained within the fiber. A laser is frequently pulse modulated to generate the optical signals, while a photo-detector converts the optical signal into an electrical signal again at the destination. The allowable pulse rate (bandwidth) of the fiber is inversely proportional to its length due to mode dispersion and pulse broadening. Pulse-rate/length products of well above 100 M bit km/-sec have been achieved. For a laser having a 1Å spectral width, the limit would be near 50,000 M bit/km-sec. Thus for a spacing of 4 miles, the limit on pulse-rate for fibers is well above 1000 M bit/sec. Second generation fibers having graded refractive indexes were producing bandwidths of 400 MHz at 1 km in 1976 (8).

The advantages of fiber optics include increased bandwidth, smaller size, less eight, higher noise immunity, and potentially lower cost when compared with coaxial cable and wire. Fiber optics has potential application everywhere thse media are used. Trunking in metropolitan areas has been the first application area, but many others will likely follow, including local connections and even connections within equipment. The high immunity from electromagnetic interference makes fiber optics attractive for applications where noise immunity is important such as in automotive and aircraft applications. While much has been accomplished in fiber-optic communications technology, much remains to be done in the areas of optical sources, detectors, repeaters, and the fibers themselves.

Applications for communications wtihin the national airspace system are sufficiently diverse that all of these competing transmission modes will be important. For ground-ground systems, fiber optics will content with waveguides, coaxial cables, and microwave radio, while for air-ground links, radio will naturally remain important, enhanced by satellite systems. On board, both fiber optics and conventional wire are available options. The forecasts will attempt to compare these various approaches and the performance levels they will acchieve during the next two decades of development.

1.3 Assumptions for the Unconstrained Forecasts

Any technology forecast involves a number of assumptions, explicit or implicit, on which the forecast is based, even if those assumptions only constitute a continuation of the climate of the recent past. It is the purpose of this section to identify the major factors which will influence the future of telecommunications and make the assumptions regarding them explicit. The discussion at this point will be largely qualitative and in some cases relative to the recent past, where the evolution of technology is now a matter of historical record.

Since for the most part telecommunications fits within the boarder sphere of electronics, many of the assumptions will be discussed with regard to this broader electronics technology on which telecommunications depends. However, in certain key areas such as regulation, major differences between electronics and telecommunications can be expected and the discussion will then deal specifically with the latter area. The specific assumptions for the forecasts involve six key areas: market demand and growth, competition in the industry, the nature of telecommunications companies, the regulatory climate, manpower availability, and the availability of needed materials. Each of these factors is discussed below.

1.3.1 Future Growth in the Telecommunications Market

The telecommunications industry will continue to experience rapid growth. Traditional markets for

> electronics such as computers and communications will expand rapidly and large new electronics markets will evolve in transportation, health care, manufacturing, and consumer products. These more sophisticated products will generate further demands for communications, reinforcing the rapidly escalatinng market. As electronics pervades more and more societal sectors, the demand for telecommunications equipment will soar. Market demand will be an increasingly powerful influence on the industry during the remainder of the twentieth century.

In the past, telecommunications has enjoyed a significant annual growth but one which is probably better characterized as moderate rather than explosive. This growth typically consisted of the introduction of a product (e.g. the telephone, television, etc.), its relatively rapid penetration into the marketplace, and finally at least partial market saturation, leaving a relatively strong and stable replacement market. In the United States, the majority of homes had telephones by 1950 and subsequent growth came through the general population increase, business expansion, extension phones, and replacements. Efforts to broaden the scope of telephone usage met with only limited success, a well-known example being the Bell System's Picturephone Service. Television began making inroads into the consumer market in the late 1940's and the black and white market had matured by 1960 when color sets made their appearance. In consumer-oriented peripherals and in telephones, the average lifetime of the equipment is rarely more than ten years and the design/development time is less than three years. This makes the peripheral/input-output area relatively fast-moving and likely to employ the latest advances in technology. The market volumes are very large.

The approximate design times and service lifetimes for input-output (peripheral), switching, and transmission systems are summarized in Figure 1.4 (9). Design times for switching devices range from about three years for small key systems and private branch exchanges (PBXs) to six or eight years for a large central office (CO) machine. Lifetimes range from fifteen years or so to forty years, which is the service goal for CO equipment. Thus, these markets move slower and require much more substantial development efforts. For a large toll switching machine, peak development costs can approach a million dollars per day.

It is perhaps useful to examine the markets for electronics and telecommunications in greater detail. In 1977, the world sales of U.S. integrated circuit manufacturers was about 2.6 billion dollars (10). This market grew to 3.2 billion in 1978 and is expected to reach 3.7 billion in 1979, an increase of about 42 percent in two years. The top 10 companies account for 78 percent of the total sales. These numbers reflect only the sales of merchant suppliers and exclude the output of the captive manufacturers (integrated circuit production facilities within a larger company which sell only to that company). Figure 1.5 (11) summarizes the world market for large-scale integrated circuits (LSI) between 1974 and 1985 in 1975 dollars by region. The LSI market grew from $0.9 billion dollars in 1974 to $1.73 billion in 1978 and is expected to reach $5.4 billion in 1985 -- an expansion of six times in 11 years. The end use markets are also shown in Figure 1.5. In 1975,

Design and Lifetime for Various Types
of Telecommunications Equipment [9]

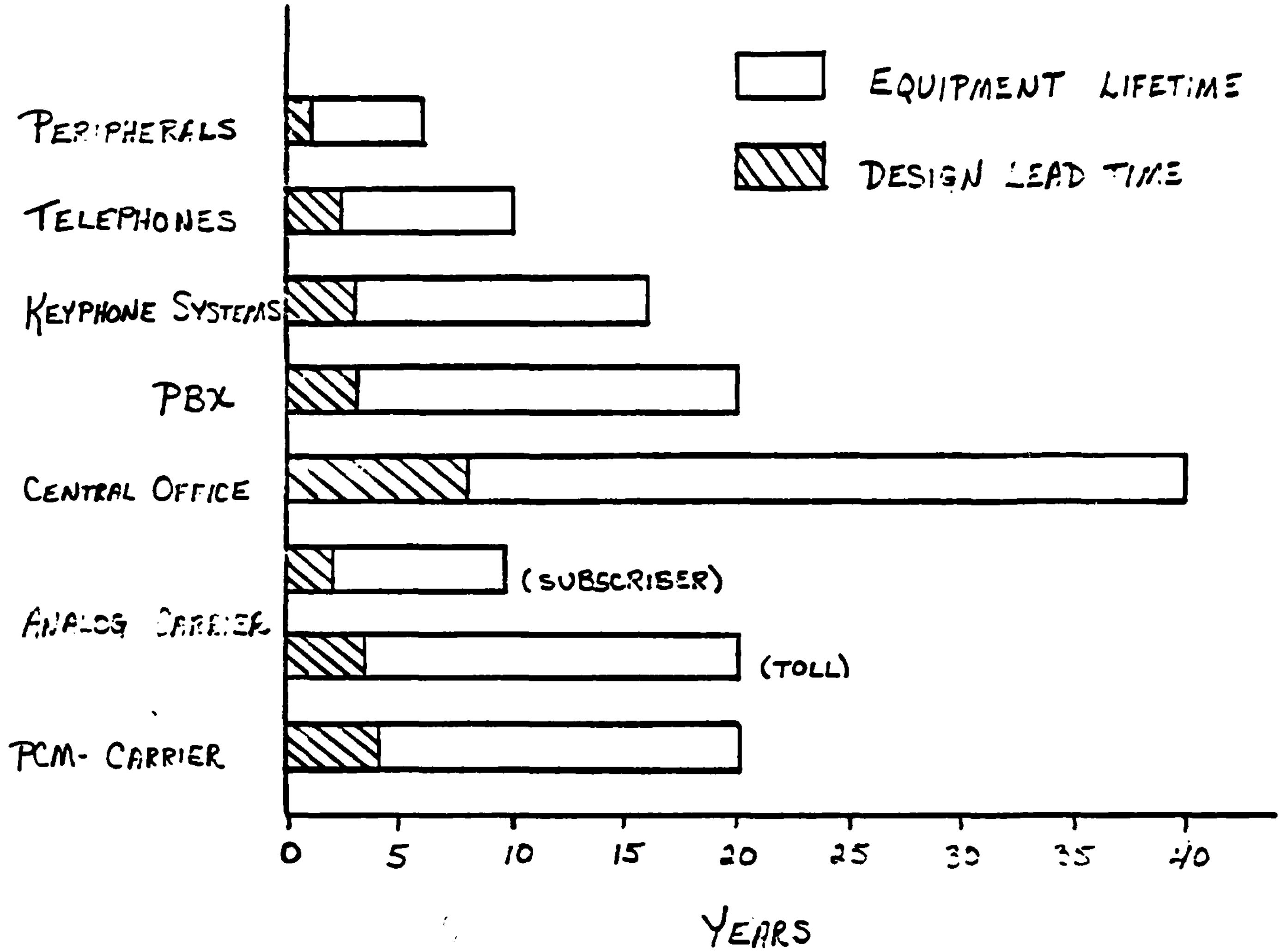


Figure 1.4

End Use Markets for LSI

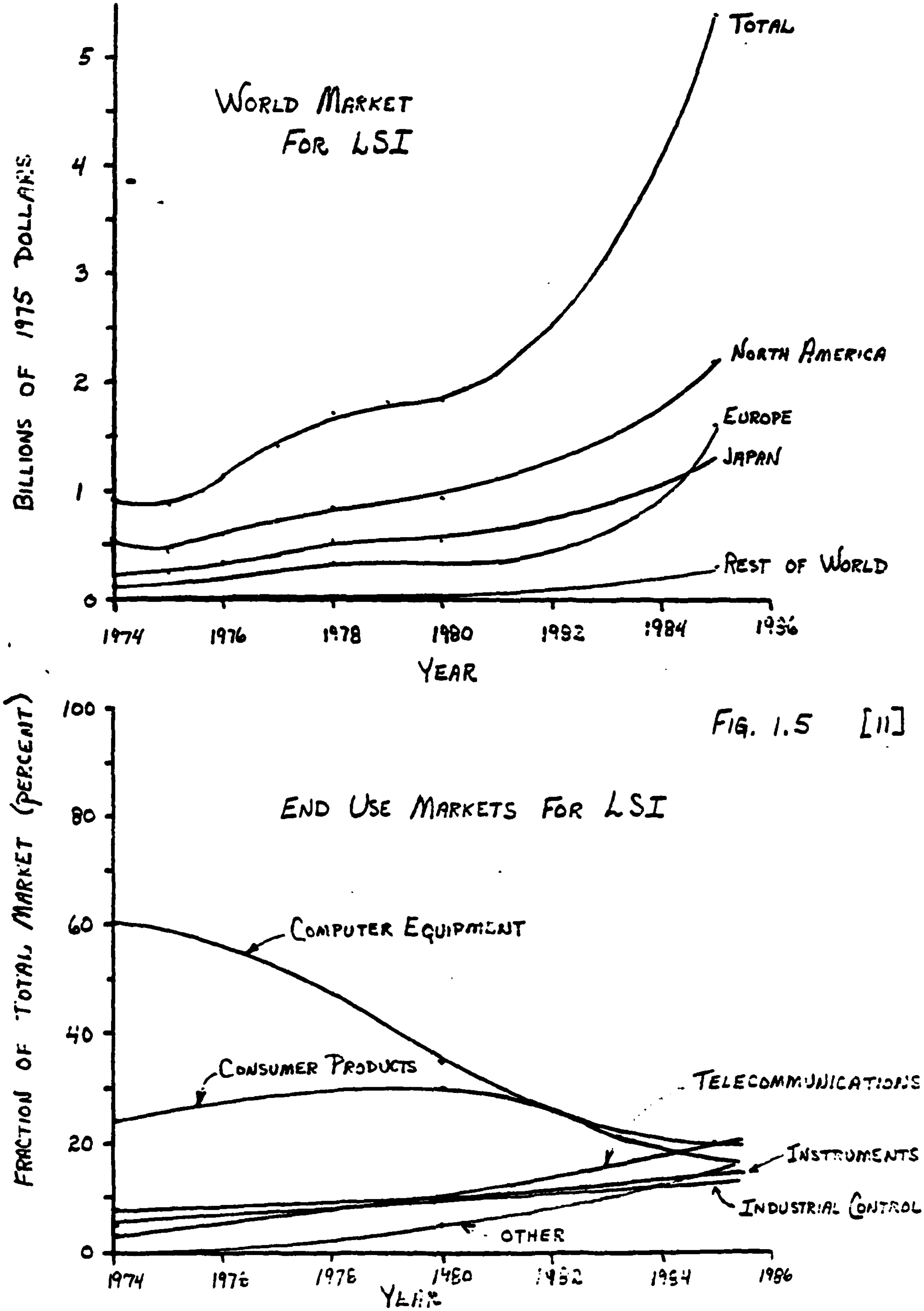


Figure 1.5

ACUMENICS

computer equipment and consumer products (calculators, watches) dominated the market at 60 percent and 24 percent, respectively. By 1985, the LSI market is much more evenly divided among six categories. The market for LSI in telecommunications is expected to grow from $26 million in 1975 to $180 million in 1980 and to more than one billion dollars in 1985. More than half of this market is expected to be in the input-output area. The total telecommunications market is, of course, much larger than these numbers indicate and involves a great deal more electronics than LSI. In 1978, the estimated world sales of telecommunications equipment was $16.7 billion dollars, with the leader, Western Electric, accounting for about $5 billion of this amount (9). The estimated semiconductor content of this equipment, including discretes, hybrids, and ICs, was nearly $1 billion. Nevertheless, it is LSI that is making the future improvements in telecommunications equipment possible.

It should be emphasized that the increasing expansion of all markets for integrated electronics and telecommunications is vital to the continued growth of the electronics industry. As pointed out by Moore (12), if the industry continues growing at its present rate, it will be producing nearly a quarter of a million circuit functions (memory bits, logic gates, etc.) per year for every person on earth in 1985. To avoid market stagnation, major new market areas are required. One area where this growth is expected is in telecommunications.

It is also expected that technology and the ability to produce low-cost sophisticated input-output terminals for the home will probably outspace our ability to apply the technology in the national telecommunications network. Limitations in available electromagnetic spectrum and the enormous task of modifying the existing outside plant to accommodate increased message rates are significant problems. Nevertheless, progress will be made and at an increasing rate.

We are now in the very early phases of the development of many new telecommunications systems. A large number of LSI chips are now in design and should be in place in telecommunications equipment during the early 1980s. However, it is during the decade from 1985 to 1995 that the largest growth in telecommunications is expected. Thus, the assumption of strong and increasing growth for telecommunications is expected to be valid throughout the rest of the twentieth century.

1.3.2 Competition in Electronics and Telecommunications

As integrated electronics and telecommunications assume an increasingly important role in society, they will become focal points for increasingly intense competition. This competition will occur not only among companies, but among nations, since supremacy in these high-technology areas will be vital to national interests. Increased competition will accelerate progress and enhance the rate at which new products are introduced.

From a historical perspective, solid-state electronics began as the result of efforts in telecommunications, and dates from the

invention of the transistor at Bell Telephone Laboratories in 1947. Most of the early development work on transistors was done at Bell Laboratories, but the adoption of solid-state devices for telecommunications equipment was relatively slow, partly because of reliability concerns. On the other hand, there were many markets outside telephone communications where reliability was less important. The transistor radio is one example. During the 1950s, several corporations such as Fairchild Semiconductor, Texas Instruments, Motorola, Philco, and others took the lead in applying solid-state devices and in improving their associated fabrication technology. A breakthrough occurred in 1959 with the development of the planar fabrication process, which permitted not only high-volume batch production of discrete transistors having improved performance but also the integration of entire circuit functions in a single slice of silicon. The decade of the 1960s saw intense efforts aimed at the development of improved fabrication techniques for integrated circuits. Diffusion, epitaxy, ion implantation, and many other processes were developed during this period. The success of these efforts in process technology were seen in dramatic increases in chip sophistication. The number of circuit elements which could be successfully integrated doubled each year and device performance improved steadily.

The decade of the 1960s saw rapid increase in the number of companies engaged in integrated circuit development. These new companies were high-technology component suppliers, and many were

spinoffs from Fairchild Semiconductor. The San Francisco Bay Area quickly became known as Silicon Valley and for good reason. Competing companies became densely packed and often occupied adjacent buildings. Job-hopping became relatively common, resulting in a relatively high turnover of senior technical personnel and relatively short lead times for any given company over any other in a given product line. Competition was intense, and developments were rapid. Some companies failed but many succeeded. By the late 1960s, few of the industry leaders of the 1950s still commanded a position of leadership in integrated electronics (13).

By the early 1970s, venture capital had tightened considerably and during the 1970s, few new integrated circuit companies were formed. However, two other developments occurred to strongly influence competition in the industry. First, by the mid 1970s, the technology had matured so sufficiently that entire systems could be integrated on a single chip. This forced many integrated circuit firms, which formerly were component suppliers, to become concerned with end products, i.e., to enter the systems world. This transition was not always easily made but many merchant suppliers became involved in marketing end products. At the same time, the technology became so pervasive due to the rapidly declining cost of sophisticated digital control, that many system houses established captive process facilities. Between 1975 and 1979, the number of merchant suppliers in the electronics industry remained virtually constant, while the number of captive suppliers

nearly doubled (14). Thus, the degree of competition for the business of the systems houses to competition for the business of the general consumer, i.e., the man on the street.

A second major change brought on by the rapid improvements in integrated circuit performance and cost was that of increased international significance. Until the 1970s, integrated electronics was dominated by American companies. It was generally recognized by the mid-1970s, however, that integrated circuits, and LSI in particular, hold the key to success in telecommunications, computers, and defense in the next century. Accordingly, integrated electronics was recognized as an arena of vital national importance by 1976 and nearly every major industrialized nation was developing a government program to assist in technology development. In Europe, major programs have been undertaken by England, France, and West Germany (15). These programs are intended to allow these countries to maintain some capabilities in very large scale integration (VLSI), and involve investments in the range of $100 million dollars over a several year period. It is not expected that Europe will achieve dominance in VLSI, although considerable cooperation within the European community is likely.

The main competition of the international level is between the United States and Japan, and here a real struggle for dominance is occurring. Since 1976, five major Japanese corporations (Fujitsu, Hitachi, Mitsubishi, Nippon Electric, and Toshiba) have

been working (with massive government support) to establish supremacy for Japan in VLSI. The size of this program has been estimated at between $200 and $400 million dollars. In the United States, this has been countered by a DOD-sponsored program at around $200 million which should begin by 1980 (16,17). During the last four years, competition between the United States and Japan has amounted to an intercontinental VLSI battle, with the U.S. maintaining a position of leadership but with that leadership being threatened, especially in the important memory area. Considerable frustration has been evident on the part of U.S. manufacturers, generated by unequal tariff laws and by the "spying" of the Japanese on U.S. firms while closing their own plants to foreign visitors.

> "Their intent is quite clear. They are out to slit our throats, and we'd better recognize that and do something about it." "...they have the unfair advantage of having their government on their side, and it seems that they have our government on their side as well."
>
> Robert N. Noyce, Chairman of the Board, Intel Corporation (18)

> "Japan does not play by the same rules that everyone else does in the international market. In the U.S., we are open for the marketing of Japanese products, but a U.S. company selling in Japan faces increased import duties, import quotas, requirements for import licenses, red tape, and the necessity to establish a presence -- but a simple presence such as a sales office takes years to establish. We also face what I call the Japan Club, an unwritten policy among Japanese users to buy outside only what cannot be obtained locally in Japan."

"To gain technology, the Japanese are establishing listening posts in the U.S. They have strict control on foreign investments in Japan to protect their own markets, control tied to the necessity for licensing arrangement. They have a patent policy that effectively forbids U.S. ownership, or non-Japanese ownership, of the most basic patents in this industry, patents on devices that were invented and paid for by U.S. industries."

Floyd Kvamme, Vice President,
National Semiconductor
(18)

"Japan virtually dominates the production of all manufactured goods. There appears to be a "conspiracy of incompetence" on the part of U.S. government, industry and academia which actually encourages the domination. While we still hold a dominant position in the production of semiconductors and computers, this position is fading. Hence, the need for goods and the money to support research (which comes from taxes on this remaining flow of goods) is dwindling. However, although at a society level, the U.S. position is completely doomed, the scientific and engineering community can still play a minor part by providing the Japanese with research, education, ideas and product prototypes. While this will be acceptable in the short term, we must worry about not getting the feedback that stimulates knowledge when we're forced to build."

Gordon Bell
Digital Equipment Corp.
(19)

Thus, competition is now intense between companies and is complicated by international policies as well. Given the national importance of VLSI, it is likely that integrated electronics and telecommunications will remain focal points of intense activity and competition well into the twenty-first century.

1.3.3 The structure of Telecommunications Companies

- The telecommunications industry will continue to be characterized by large, highly-structured companies having the resources to tackle large system developments. These companies will be somewhat more responsive to market pressures than in the past and will have somewhat less internal inertia. Companies involved in peripheral devices will be an exception. These firms will be smaller, faster moving, and more entrepreneurial in nature and will exist in greater numbers.

In the past, telecommunications has been characterized by very large companies. AT and T (including Western Electric), ITT, GTE, Nippon Electric, Northern Telecom, and Siemens are a few examples. This situation was almost mandated by the size of the development efforts required. Something akin to territorial sovereignty was also a result and vigorous competition among many companies was muted if not completely suppressed. With the increased use of integrated electronics, fiber optics, and other new technology, the required development time for a given level of performance is shrinking, however, and many development jobs are becoming tractable in more companies as a result. The boundaries between telecommunications and other areas are also becoming increasingly uncertain. With computer-controlled digital networks switching digital bit streams representing voice, data, text, and images, the nature of telecommunications has broadened sufficiently to overlap areas served by corporations such as IBM and Xerox. It is likely that this overlap will grow, resulting in increased competition in both the transmission and switching areas. Combined with increased

market demand for more service and new features, the major companies are expected to move faster, be more sensitive to market needs and be more aggressive in market policy. Large undeveloped overseas markets should play a strong role toward the end of the twentieth century as advanced telecommunications becomes worldwide.

In the area of peripheral devices, things will move faster and be fiercely competitive. The sophistication of these devices will grow to encompass two-way voice, video, and facsimile reproduction. Development times will remain relatively short, however, so that the area will remain attractive to relatively small companies. The terminal area will involve present manufacturers of television, printers, computer terminals, telephones, and copy machines as well as new companies seeking to meet the market demand. While not all will be successful, it is likely that several dozen major manufacturers of terminal equipment will compete for the world-wide market during the 1990s.

> 1.3.4 <u>The Regulatory Climate and Its Impact</u>
>
> The regulatory climate will relax from past levels in response to market demand and the broadening scope of telecommunications to allow increased competition and more rapid growth. Standardized bus structures and protocols will be established to permit the interconnection of equipment from different manufacturers. Tariffs and foreign policy will continue to restrict, but will not prohibit, major market areas.

One of the major differences between the electronics industry and the telecommunications industry has occurred in the area of

regulation. While the electronics industry has been largely unregulated, telecommunications has been highly regulated. This regulation has ensured a well-operating, orderly communications network, but in some areas has probably slowed progress. During the past decade, the degree of regulation has relaxed somewhat, particularly in the terminal area (input-output devices). It is expected that this trend will continue as attempts are made to simultaneously preserve the benefits of regulation while encouraging increased competition and more rapid progress. It is assumed that these efforts will be largely successful, and that while the switching and transmission areas, and the local plant, continue to see relatively strong regulation, the input-output area will become largely unregulated with the exception of standardized protocols which will govern interconnections between equipment.

Telecommunications is expected to be an increasingly high-technology business, with basic equipment manufacturing controlled by relatively few companies in the U.S., Japan, and Europe. While sovereignty will likely mandate local operation of national telecommunications networks, most developing countries are expected to buy rather than make the basic equipment in order to benefit from advances in technology and new features. Tariffs will continue to restrict markets within producing countries to some degree, but major markets within the developing nations will serve as a primary attraction to all manufacturers by the end of this century.

1.3.5 Availability of Trained Personnel

> As integrated electronics and telecommunications are applied in more and more areas, the demand for engineers in these areas will grow. This demand will be only partly met by universities, and progress in some areas will be slowed by the lack of sufficient numbers of design engineers.

The United States has acted for many years as educator to the world, particularly in solid-state electronics. During the 1960s, large numbers of Japanese students flocked to American universities to earn degress in electrical engineering and learn solid-state technology. Since Japanese technology has become well established in its own right, they have virtually disappeared from American campuses to be replaced by students from Korea, Taiwan, and the Middle East. Proper preparation for a career in solid-state electronics requires at least a Master's Degree and in many cases a Doctorate. In the late 1960s, graduate education was very popular in the United States and graduate programs were relatively at many schools. However, graduate programs suffered during the recession of 1970-72 and never fully recovered. Many programs are now struggling at a fraction of their former capacity. Many American students are now going to industry on completion of the Bachelor's Degree, lured by relatively high pay and good facilities. The result has been graduate programs of decreased size, populated in large part (as high 50 percent at some schools) by foreign students. Since many of these students return home after graduation, the supply of trained personnel to American companies is further diminished and foreign technology is enhanced.

In addition to the considerations mentioned above, several other factors have significantly impacted the supply of trained personnel to the industry. The rapid progress and increased sophistication of integrated electronics has created an area of high technology requiring major equipment investments and a subsequent high level of support in terms of material supplies as well as staff. Many items of process equipment now cost in excess of $150,000. As a result, very few universities have found it possible to maintain a quality graduate program in the solid-state area. There is probably no more than a half dozen universities in the United States where state-of-the-art doctoral research in integrated electronics is possible. Thus, while many schools may provide an adequate undergraduate education, relatively few are adequate at the graduate level in the solid-state area. In other disciplines relative to telecommunications, such as computers and electromagnetic fields (transmission), the situation is less equipment dependent but has nonetheless suffered due to the tight budgetary conditions at most schools.

The serious situation regarding the supply of trained personnel in solid-state electronics is now being recognized and has recently been the subject of considerable discussion. It has been proposed as the subject of a special session at the 1980 International Solid-State Circuits Conference. In the United States, the solution most likely will require the combined efforts of universities, industry, and the government. Combined efforts of this sort have

already occurred in some countries. U.S. universities have interacted strongly with the federal government for many years and it is government support of university research that has permitted the present facilities to exist. However, government must be increasingly sensitive to the problems faced by universities and to the national importance of solving them. While industry has been the prime user of the universities' most important product, there has been relatively little assistance given to universities by American industry. What assistance has been given has typically been on an individual company-to-university basis and usually either in the form of used equipment or as a gift to the university as a whole, which has little effect on a particular program. On the other hand, few university programs have approached industry with a reasonable plan for collective support.

The main effort to achieve excellence in graduate engineering education and meet the manpower needs of industry must come from the universities, with funding derived from industry and governmental sources. The development of regional centers of excellence is an essential and reasonable goal for the United States. Between eight and ten centers of graduate research and education, each providing facilities for other universities within its region, would not only allow needs for trained manpower to be met but would provide a valuable source of research and development to the industry. One such center was recently established by the National Science Foundation at Cornell University.

It is expected that the United States will move slowly in the creation of regional centers and that progress will depend most strongly on the aggressiveness of particular schools. Between five and ten universities will likely maintain quality graduate programs, while the rest will be unable to keep up with the rapid pace of technology, particularly in areas requiring hardware and experimental research. University-sponsored continuing education programs as well as internal industrial efforts at on-the-job training and retraining will help, but manpower needs to support the growing telecommunications revolution are expected to continue to outpace supply. This will likely reduce the rate of progress that would otherwise be achieved and cause some projects to be deferred.

1.3.6 Availability of Material Resources

> Material resources to support the growth of telecommunications will be available in adequate amounts, and shortages will not retard progress.

The size of the telecommunications industry and sheer numbers associated with most products make the continuing availability of material resources an important concern. Fluctuations in the price and supply of copper in 1973, for example, caused significant concern within the Bell System, which uses more than half the copper produced in the United States. It is also true that the supply of silicon wafers has been very short during 1978-80. However, new production facilities are coming on line and since silicon is one of the most abundant elements on earth, no serious long-term

shortages are expected. The trend toward fiber optics replacing copper wire should also ease the copper requirements associated with new outside plants and move toward more abundant sources. The energy requirements of future operating equipment are also expected to decline. While a variety of chemicals are used in producing integrated circuits, processing fiber optic cable, and building telecommunications equipment in general, it is unlikely any serious shortages will occur. While this situation should be monitored closely, material resources are expected to be adequate to support the growth of the industry.

1. C. E. Shannon, The Mathematical Theory of Communication. Urbana, Illinois: University of Illinois Press, 1948.

2. T. H. Crowley, G. G. Harris, S. E. Miller, J. R. Pierce, and J. P. Runyon, Modern Communications. New York: Columbia University Press, 1962.

3. L. W. Roberts, G. E. Lundquist, and D. E. Findley, "The Role of Communication Systems in Air Traffic Management," IEEE Trans. Communications, 21, pp. 346-363, May, 1973.

4. H. Falk, "Telephone Technology: The 'Which Switch' Game," IEEE Spectrum, pp. 41-42, June, 1975.

5. A. E. Joel, Jr., "ESS: Minimonster," IEEE Spectrum, pp. 33-35, August, 1976.

6. S. E. Scrupski, "Users Starting To Hop Aboard U.S. Communications Satellites," Electronics, 47, pp. 95-102, October 3, 1974.

7. I. Jacobs and S. E. Miller, "Optical Transmission of Voice and Data," IEEE Spectrum, pp. 33-41, February, 1977.

8. R. Gundlach, "Fiber-optic Developments Spark World-wide Interest," Electronics, pp. 81-104, August 5, 1976.

9. R. Grundlach, "Large-Scale Integration is Ready to Answer to Call of Telecommunications," Electronics, 50, pp. 93-108, April 28, 1977.

10. ICE Corporation Staff, "Status 1979: A Report on the Integrated Circuit Industry," Integrated Circuit Engineering Corporation, Scottsdale, Arizona, 1979.

11. SRI Staff, "LSI Circuits," Long Range Planning Report 583, Stanford Research Institute, December, 1976.

12. N. Lindgren, "Semiconductors Face the '80s," IEEE Spectrum, October, 1977, pp. 42-48.

13. I. M. Mackintosh, "Large Scale Integration: Intercontinental Aspects," IEEE Spectrum, oo, 51-56, June, 1978.

14. G. Moore, "Are We Really Ready for VLSI," Keynote Address, 1979 International Solid State Circuits Conference, Philadelphia, February, 1979.

15. U. K. National Economic Development Office Report, "The Microelectronics Industry," in Microelectronics 9, pp. 37-42. September, 1978.

16. Science Staff, "Microelectronics: Defense Department Looks to the 1980's," Science 201, pp. 1112-1113, September, 1978.

17. Staff Report, "VHSIC Planning Progresses," Electronic Weapons/Defense Electronics, pp. 36-38, April, 1979.

18. G. Bylinsky, "The Japanese Spies in Silicon Valley, Fortune, pp. 74-78, February 27, 1978.

19. G. Bell, "Limits in the Opportunities for Using Microstructures in Computer Systems and Architecture," 1978 Arlie Conference Proceedings.

Chapter Two

Microelectronics Technology

2. An Unconstrained Forecast of Microelectronics Technology

As indicated in Section 1 of this report, telecommunications represents, for the most part, an important application area within the broader sphere of solid-state electronics. As electronic functions become more highly integrated and as the cost of electronic data processing and control decreases, telecommunications is expected to become still more closely associated with microelectronics and solid-state technology, particularly in the input-output and switching areas. In the transmission area, telecommunications needs are expected to lead electronics into new areas, especially in optics and optoelectronic devices.

Since developments in microelectronics are key in forecasting progress in many areas of telecommunications, this section examines microelectronics in greater detail and presents an unconstrained forecast of microelectronics technology. This forecast is subject to the general assumptions discussed in Section 1.3. In many areas, the discussion is drawn from an earlier study (1), abbreviated somewhat and updated to reflect recent progress and present views of future developments.

2.1 General Comments on Integrated Elecronics

Solid-state electronics began with the invention of the transitor at Bell Telephone Laboratories in 1947. During the 1950's the bipolar transistor structure was refined and transistorized circuits

rapidly replaced vacuum tubes in a wide variety of applications. About 1960, batch-fabrication and miniaturization were extended to the circuit level with the development of the planar process for integrated circuits. This process was based on the use of photo-engraving to define transistors, resistors, and capacitors in a single monolithic silicon substrate. The 1960s concentrated on the development of technology. The art of photolithography was extended, ion implantation, epitaxy and diffusion were introduced and refined for selectively introducing needed elements into the silicon lattice, and a variety of metal interconnect schemes were investigated.

Three driving factors motivated these developments in technology: achieving lower cost, higher speed, and higher precision. The drive toward lower cost concentrated on reducing device size and on simplifying the fabrication sequence to put more functions on a chip and attain higher yields. During the first half of the 1960s, the technology rested entirely on the junction-isolated bipolar transistor. About 1965, metal-oxide-silicon (MOS) transistor circuits were introduced for digital applications. Although slower than bipolar circuits, they were also more dense and more easily fabricated. Since the mid-60s, bipolar - MOS competition has been increasingly intense. Bipolar technology has become more dense while MOS devices have become faster. By the mid-1970s, MOS technology dominated all digital applications except those requiring

extremely high speed, and was making some inroads into analog functions as well. Many variations of the basic device approaches were developed and at least five were competing for LSI applications in 1980. The diversity of approaches and the number of companies involved in developing integrated electronics had a strong influence on the rapid pace with which developments occurred.

Although most applications were digital, from 1965 onward an increasing number of analog integrated circuits were introduced. Operational amplifiers were the first major analog circuits to be realized in monolithic form. Later, phase-locked loops, digital-to-analog and analog-to-digital converters followed. These developments illustrated efforts aimed at achieving high precision from a batch process, including low temperature and power supply sensitivity and high accuracy. Many structures at first used precision thin-film passive components and silicon active devices in hybrid assemblies; however, the trend has been increasingly toward monolithic structures for lower cost while preserving performance by sophisticated circuit design, and, in some cases, by the use of thin-film components deposited directly on the silicon chip itself.

Figure 2.1 shows the ability to pack electronic components on a single chip (die) while maintaining acceptable yields. The number of components per chip has been doubling every year and in the near term is expected to continue doubling every two years or less. This trend has come to be known as Moore's Law, after Gordon E. Moore of Intel Corporation who first pointed it out.

Trends in the Cost and Level of Integration
In Monolithic Circuits

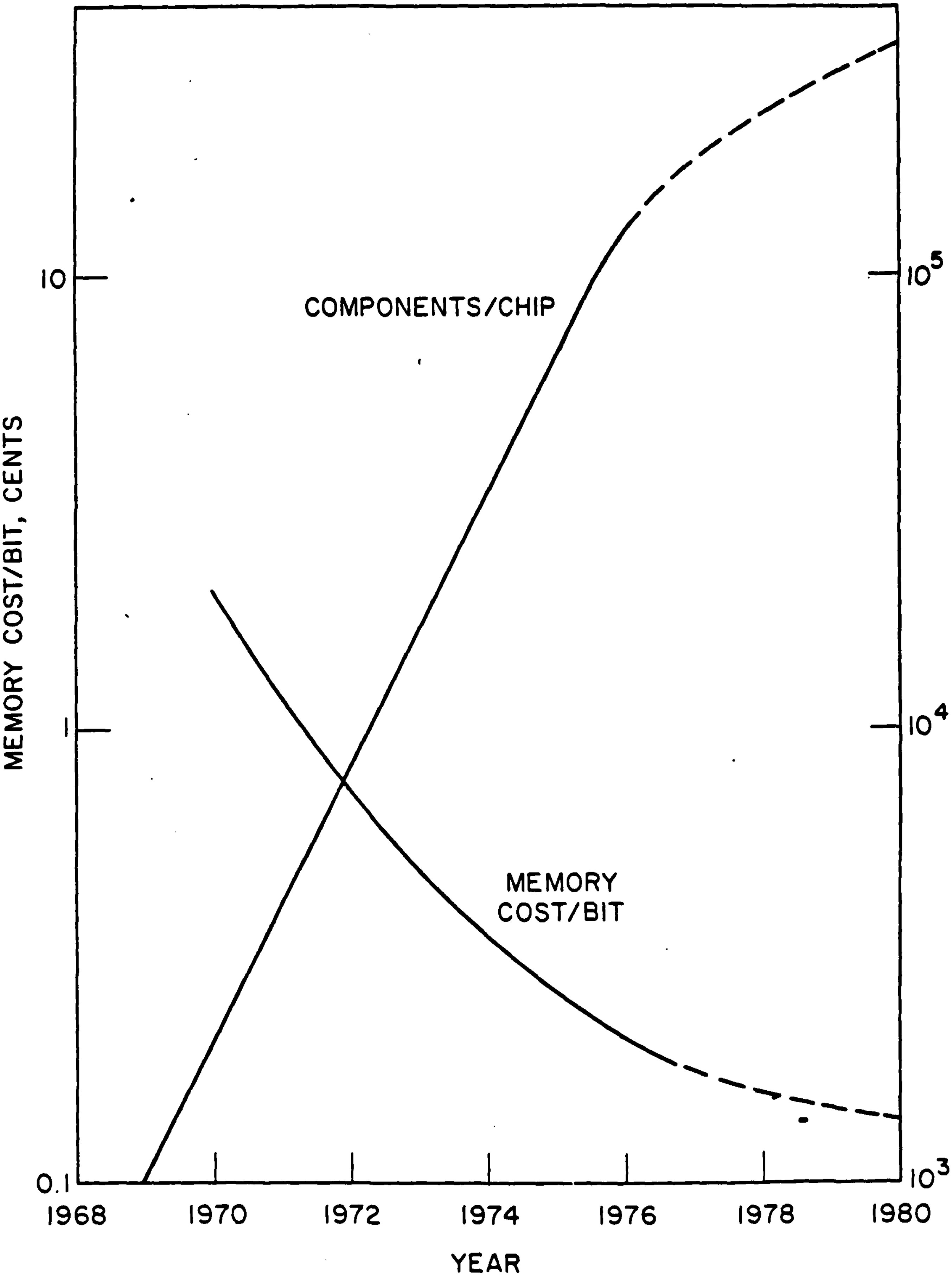


Figure 2.1

ACUMENICS

At the same time, the cost per function (in this case, per bit of memory) has been declining rapidly. It is this combination of increased functional complexity and decreased functional cost which is causing the rapid proliferation of electronics into many formerly nonelectronic areas, as well as rapid progress in many traditionally electronic areas such as telecommunications.

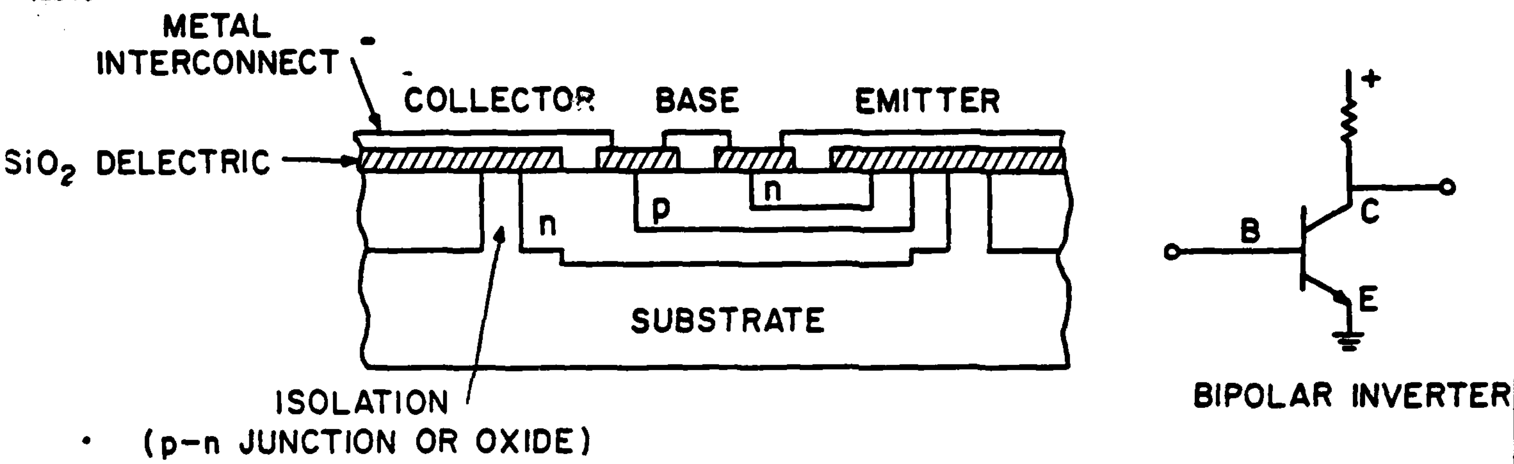
From an applications standpoint, most computer designs during the 1950s and 60s were oriented toward the development of large, high-speed, highly centralized computers in applications ranging from general computation to telephone switching. The impact on society in general was significant but subtle. By the early 1970s, integrated electronics had developed to the point where the number of electronic components which could be integrated on a single chip of silicon was sufficient to allow complete processors for calculators and small computers to be integrated in an area typically three to five millimeters on a side. The term "microprocessor" was quickly applied to these new devices, and as technology evolved still further, permitting significant amounts of memory to be included on the processor chip, the term "microcomputer" was applied to these still more densely integrated systems.

The introduction of microprocessors and microcomputers represents an important turning point in the development of electronic systems. The availability of extremely small, inexpensive computers having

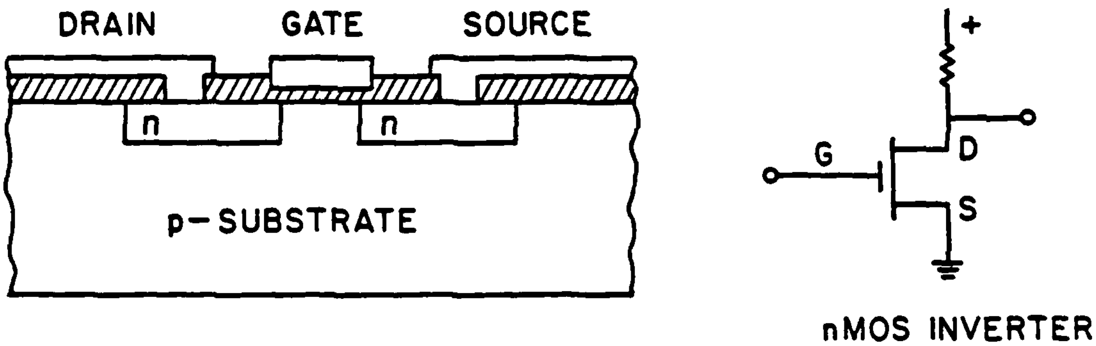
significant processing power and versatility has allowed stored-program control and data processing to be extended into even the smallest electronic instruments. Distributed, decentralized computer control and the digital handling of information are now becoming widespread -- one of the best examples being in telecommunications.

2.1.1 Basic Device Technologies for Large Scale Integration

Present integrated circuits are formed via a series of batch photoengraving operations which control the selective introduction of dopants such as boron (p-type) and phosphorous (n-type) into the silicon lattice to form diodes, resistors and transistors. The implementation of logic with these components depends on the realization of an *electronic switch* (e.g., a device in which the application of a small voltage or current at one terminal controls the flow of a larger current between two other terminals). For the last ten years, the two ways of implementing this switch in silicon have competed. Figure 2.2 shows cross sections of the two basic approaches: the charge-controlled bipolar transistor and the voltage-controlled unipolar MOS transistor. The bipolar transistor bases its operation on the fact that a positive voltage applied to the base relative to the emitter causes electrons to be injected from the heavily-doped emitter into the base. Once injected, the carriers diffuse to the reverse biased collector where they constitute collector current.



BIPOLAR (npn) TRANSISTOR



MOS TRANSISTOR

Figure 2.2

The only charge which flows in the external base lead is that necessary to make up for electrons which recombine before they reach the collector. Thus, a relatively small base current can control a large collector-emitter current. The on-resistance from collector to emitter when used as a switch is typically less than 100 ohms while the off-resistance is many megohms.

The MOS switch is based on a different phenomenon. With no voltage applied to the gate, no conducting path from drain to source exists and the switch is off. When a positive voltage is applied to the gate, however, holes are repelled from the silicon surface under the gate and the electrons are attracted. When the gate voltage rises above a threshold determined by the channel doping and gate dielectric thickness, the silicon surface under the gate electrode inverts to n-type, forming a conducting path (channel) from drain to source with an on-resistance from several hundred to several thousand ohms depending on the applied gate voltage and the size of the device. Details of device fabrication, design, and operation can be found in the literature (2-5).

Figure 2.2 also shows a basic inverter implemented with bipolar and MOS switches. With a positive voltage applied to the base or gate, the switch is on and the output is near ground. With an input voltage near ground, the switch is off and the output is high. The LSI technologies differ in the device structures used for the driver (switch) device and the load. Many technologies use

a second transistor as an active load device, both because the transistor requires less area than would a diffused resistor of comparable value and because the nonlinear properties of the active load can sometimes be used to improve the power and speed of the gate. In integrated injection logic (I^2L) (6), the most promising of the bipolar approaches, a lateral pnp transistor is used as a load. In p-channel and n-channel MOS, both enhancement and depletion loads have been used (4), while in complementary MOS (CMOS) the load is a pMOS device and the driver is nMOS. Since in CMOS either the driver or the load transistor is always off, there is no quiescent current path from the supply to ground and the quiescent power dissipation of this technology is virtually negligible.

All of these basic device technologies are evolving rapidly as the result of improved processes, structures, and photolithographic tolerances. The choice of a technology usually depends on the specific requirements of a given application, but an increasing number of applications recently have adopted the MOS approach, which allows high levels of integration and increasingly high speed. Comments on the future of specific technologies can be found elsewhere (1, 6-9). For the purposes of this study, it will be sufficient to examine the performance which will likely be available from any of the technologies as of a given point in time.

2.1.2 Cost in LSI Circuits

The objective cost of an integrated circuit is composed of the cost of the processed die together with the cost of packaging and testing the completed unit. The cost of processing a silicon wafer is relatively independent of wafer size and a present cost of $40 per wafer is assumed. (In these and other projections, 1975 dollars are assumed unless otherwise specified.) The number of dies per wafer, neglecting edge drop out, is

$$N = \frac{\text{wafer area}}{\text{die area}} = 0.79 \left(\frac{D^2}{A}\right) \qquad (1)$$

where D is the wafer diameter and A is the die area. The primary yield limiting mechanism is the presence of defects on the masks (or in the photoresist) which limit the yield from each photo-engraving operation. The approximate yield of die which pass visual inspection, neglecting die breakage or loss due to non-mask related sources is (10)

$$y_p = (1 + FA)^{-n} , \qquad (2)$$

where F is the defect density per mask and n is the number of masking operations per wafer. The cost of a potentially good die is

$$\text{cost/die} = \frac{\$30}{NY_p} . \qquad (3)$$

To this cost must be added the cost of testing, T, and packaging, P. A cost of \$0.30/die for testing and a package cost of \$0.15 for plastic or \$0.82 for ceramic is assumed. A final test yield, Y_F, of 0.8 is also assumed. Then, the approximate cost per packaged circuit is

$$\text{cost/DIP} = \left[\frac{\$30}{NY_p} + T + P \right] \left[\frac{1}{Y_F} \right] . \qquad (4)$$

Figure 2.3 shows cost as a function of die area with the process complexity (number of masking steps) as a parameter. For 3 inch wafers and 10 defects per square inch, the final circuit cost is dominated by packaging and testing below die areas of approximately twenty thousand square mils (20 K mil^2 = 12.5 mm^2). Most present random access memories (RAMs) fall in the 20 to 30 K mil^2 range so that assembly, packaging and testing can be expected to play an important role in the final cost of these units. Many microcomputers are in the 30 to 50 K mil^2 range, where die cost and process complexity remain more important. Figure 2.3 also shows the estimated cost of a single-chip function requring 20 K mil^2 of silicon in nMOS, assuming a 50 percent overhead at this density for lead routing. The approximate cost in other technologies is also shown.

Circuit Cost as a Function of Die Area

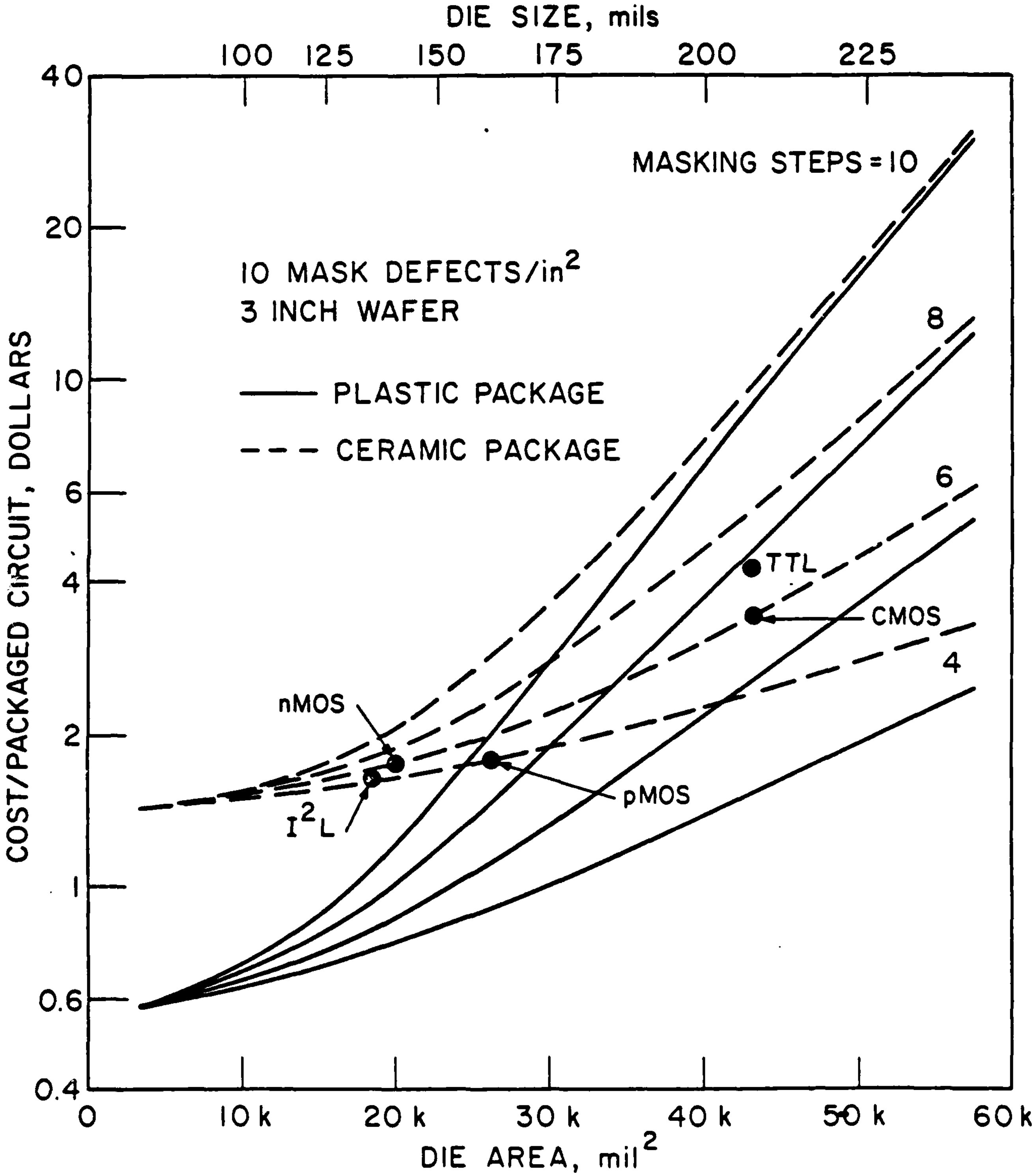


Figure 2.3

Figure 2.4 shows the packaged circuit cost per unit area as a function of die size. A 3 inch diameter wafer, 6 mask process, 10 mask defects per square inch, and a ceramic package is assumed nominal, and single deviations are examined in terms of their effect on cost. For large systems, it is clearly more cost effective to implement the required functions in a few large die rather than in many smaller chips. For example, a system requiring 40 k mil^2 of silicon in a given technology would cost approximately $3 as a single LSI chip. As four 10 K mil^2 chips, the system cost would be $6. Recent activity in projection printing (to achieve lower defect densities) and the trend to larger wafers support the use of larger, more complex chips in future LSI systems, since both produce a significant shift in the minimum cost point in Figure 2.4 to larger die sizes. The objective costs given in these figures indicate that given sufficient time to fine tune a product line, the cost of such systems will be extremely low. Microprocessors are available today at less than $5 in volume.

2.1.3 Integrated Circuit Reliability

High reliability is essential in many applications of electronics, especially in aviation and telecommunications. In addition to extensive work to minimize failures at the chip level, work on applying redundancy through the use of distributed and duplicated control is one approach being investigated to ensure high system reliability in the face of imperfect reliability at the chip level.

Packaged Circuit Cost Per Unit Area as a Function of Die Size

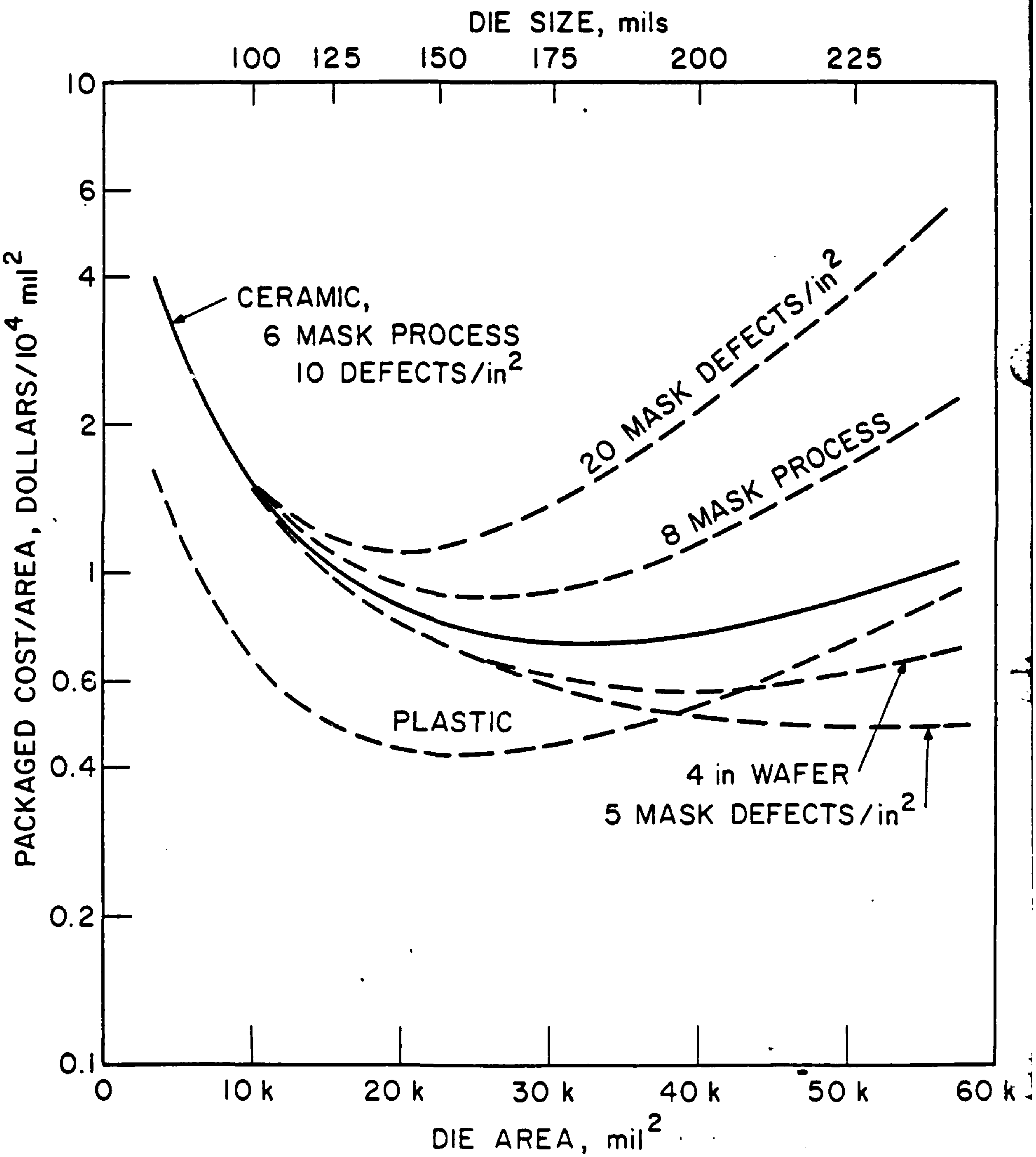


Figure 2.4

ACUMENICS

In the past, on-chip failures in integrated electronics have been frequently of less significance than off-chip failures, e.g., at bonds or associated with external leads (11). As systems become more and more densely integrated, system reliability should improve substantially, approaching the reliability level of a single VLSI chip. Manufacturer reported failure rates for both MOS and bipolar circuits, including dynamic MOS RAMs, are now well under 0.1 percent/1000 hours at 70°C and 90 percent confidence level (12). Stated differently, 90 percent of such components would survive for more than 100 years at 70°C. Details concerning extrapolations based on accelerated aging are given in the literature. The significance of testing in achieving high reliabilty and its significance in determining chip cost should be emphasized.

As individual devices become smaller and more densely integrated, most failure mechanisms become at least no worse and a few become better in their contributions to component reliability. However, there is at least one class of failures which becomes steadily worse as device size and current levels decrease. These failures are "soft" in the sense that they are temporary. They are related to the extremely low current levels and finite number of electrons associated with holding internal voltage levels in densely integrated memory and logic cells. The collision of high-energy alpha particles, released through the radiative decay of trace radioactive elements

in the chip package, with the silicon lattice provides the lattice with enough local energy to generate substantial numbers of electrons and holes. These carriers diffuse to reverse-biased junctions where they constitute enough charge and current to disrupt the electrical integrity of the cell. Since such alpha particle hits are random and since the problem grows steadily worse as the integration level increases, it is likely that these problems will force a relatively early adoption of on-chip redundancy and error detection/correction circuitry. Such design procedures should be widespread by the mid-1980s.

2.1.4 Speed and Power Considerations

Speed and power can be treated together since in both MOS and bipolar circuits, high speed can typically be traded for low power and vice versa. The power-speed (power-delay) product is a widely used figure of merit in comparing different technologies.

Speed in either bipolar or MOS technology is related to the ability of the switching device to discharge node capacitances quickly. The signal delay in passing through a logic stage is proportional to the output node capacitance divided by the transconductance of the switch. The change in output current for a given change in input voltage (the transconductance) for a high-frequency bipolar transistor is in turn proportional to the output

current level while for an MOS device, the transconductance is proportional to the square root of the output current. These different dependencies on operating current level generally give bipolar devices a speed advantage and a higher off chip drive capability when compared with MOS, although at low current levels the speed differential is less and in some cases shows signs of reversing to give the advantage to MOS.

As the size of LSI devices and their interconnections has been reduced, node capacitances have decreased as expected. Dielectric thicknesses have changed relatively little so that the capacitance per unit areas has remained relatively constant. MOS devices continue to exhibit increased speed as dimensions decrease and become submicron. For bipolar devices, a direct tradeoff between power and speed occurs only when the emitter time constant (an extrinsic delay) dominates over the intrinsic space-charge transit-time delay in the collector and the minority carrier diffusion time in the thin base region. As device size decreases, node capacitance also decreases, reducing the extrinsic emitter delay at a fixed current (power) level per gate. With future size reductions the intrinsic base transit delay may therefore become relatively more important in setting a limit on circuit speed and may further reduce the speed advantage of the bipolar technologies relative to MOS.

The power of a gate is proportional to the operating current level at a fixed supply voltage. Hence, there is a more or less direct tradeoff between power and speed in both MOS and bipolar technologies, and the power-delay product serves as a convenient figure of merit in comparing different approaches.

2.2 Gate-Level Performance in VLSI Circuits

The preceding section has briefly reviewed the historical development of semiconductor technology and made general comments on some of the factors contributing to circuit cost, reliability, power, and speed. The assumptions on which the unconstrained forecasts are based were discussed in Section 1.3. In this section, forecasts of cost, speed, power, and reliability will be given. The first three parameters will be treated at the gate level, while reliability will refer to the entire LSI chip. In dealing with cost, 1975 dollars are used throughout. The parameters are forecast in terms of the performance achievable with <u>any</u> LSI technology since in terms of the user, the technology is relatively unimportant except in terms of the performance levels achieved.

As stated earlier, these forecasts are based on an earlier study (1), and many of the forecast curves are taken from that work. Since those projections were made in 1976, it will be instructive to re-examine these forecasts four years later to see how closely they have been fulfilled. Comments on the presently-conceived future of the technology relative to the original projections will also be given.

2.2.1 Cost Per Packaged Gate

Over the past decade, the binary logic gate has been the basic building block of digital circuits. As density has increased, gates have become increasingly difficult to identify physically, but have remained the basic circuit entity. The cost per packaged gate is related to the cost per unit area of packaged silicon as well as the gate packing density. It is assumed that the system to be integrated is large enough that optimum chip sizes (lowest cost per unit area) can be used.

The cost per unit area of packaged processed silicon is composed of the cost of the processed die (silicon), the testing cost, and the cost of packaging. For a given die size, die cost is expected to decline slowly in the future, partly due to increased wafer sizes. It is assumed that the chip die size is at the most cost effective point and that this size will increase from 20 K square mils (1975) to at least 80 K square mils in 1990. The cost of fabricating a monolithic circuit will benefit from increased process automation but will nonetheless increase due to more expensive equipment and supplies. Wafer process cost is expected to increase from $30 in 1975 to $50 in 1990 as wafer size increases from 3 inches (1975) to 4 inches (1980) and 5 inches by 1990. The use of improved masks, projection printing, and later, x-ray lithography (13-15) will reduce defect densities, maintaining yield in the 20 to 30 percent range and allowing the cost per unit area of processed

silicon to decrease from about 13 cents (per 10 K square mils) in 1975 to around 9 cents in 1990. It is noted that electron-beam lithography (13-15), while permitting significant reductions in minimum dimensions by 1980, is a relatively expensive exposure technique so that its use is expected to increase the cost per unit area of processed silicon. It will be widely used for mask fabrication, however, and for some direct wafer exposure between 1978 and the introduction of practical x-ray systems.

Increased testing automation and the inclusion of diagnostic circuitry on the chip will only partially offset the increased cost of testing an increasingly complex die. An increase in testing cost from 30 cents (1975) to about 75 cents (1990) is expected for random access memory (RAM). The testing cost for more complex units (e.g., microcomputers) will be even higher.

Improvements in package forming will act to balance some of the expected increases in materials costs. Improved plastics and molding techniques will permit increased use of plastic packages with a shift from metal-lid ceramic to plastic and "cerdip" packages. Plastic packaging costs are expected to increase from 15 cents (1975) to around 35 cents by 1990.

The cost per 10 K square mils of processed, packaged silicon is expected to decline slowly from about 44 cents (1975) to around 28 cents in 1990. As was the case from 1965 to 1975, the most dramatic reductions in functional cost will come from increased functional density and larger chip size.

The density of logic gates is dependent on lithographic dimensions, device structures, and operating depletion layer thicknesses in the substrate material. Production line widths have gradually decreased from 10-12μm in the late 1960s to 5-6μm in the mid 1970s, and in 1979 were in the 3-5μm range. Hard contact printing has given way to soft contact systems, and projection printing is gaining rapidly in commercial acceptance. Electron-beam and x-ray systems for device lithography are both under exploratory development and by 1985 should reduce production line widths to 1-1.5μm.

There have been several recent studies of the fundamental limits of microelectronics in terms of size, speed, and power (16-22). For an MOS device, the basic limit on device size arises as the result of an oxide field limitation (16). The minimum size for an MOS device has been postulated to be about 1.2μm on a side with a gate oxide thickness of 0.014μm and a substrate doping level of 2.7×10^{17} cm^{-3} (16). For a bipolar device (17), the minimum device size is determined when the collector avalanche and punch-through voltages approach the desired operating voltage for the circuit. The minimum theoretical base width is about 0.08μm with a corresponding device size of about 1.8μm per side. For both bipolar and MOS circuits, allowable power dissipation may set a more severe limit on packing density than does the basic device, depending on the details of circuit operation.

Figure 2.5 gives a forecast of gate density (area) over the rest of this century. The data includes an additional fifty percent for lead routing. In 1976 the technologies producing the smallest area (I^2L and nMOS) were yielding gate sizes of about 5 to 6 square mils. Since that time few real structural changes in the technologies have occurred, but gate sizes have nonetheless decreased. Improved lithographic tolerances have enabled significant downward scaling of device sizes, and the development of miltiple-level interconnect schemes (e.g., two levels of polysilicon plus one level of metal) has further reduced the space requirements for random logic. An area reduction of between two and three during the past four years from these sources is estimated, putting the present gate size between two and three square mils, which is in good agreement with the forecast.

There has been little discussion during the last few years concerning non-binary (higher radix) logic and it remains uncertain whether these approaches will ever be used in production components. However, they remain a possibility for the 1990s, when binary bates are expected to approach presently-conceived limits on device size.

Forecast of Gate Density in LSI Circuits

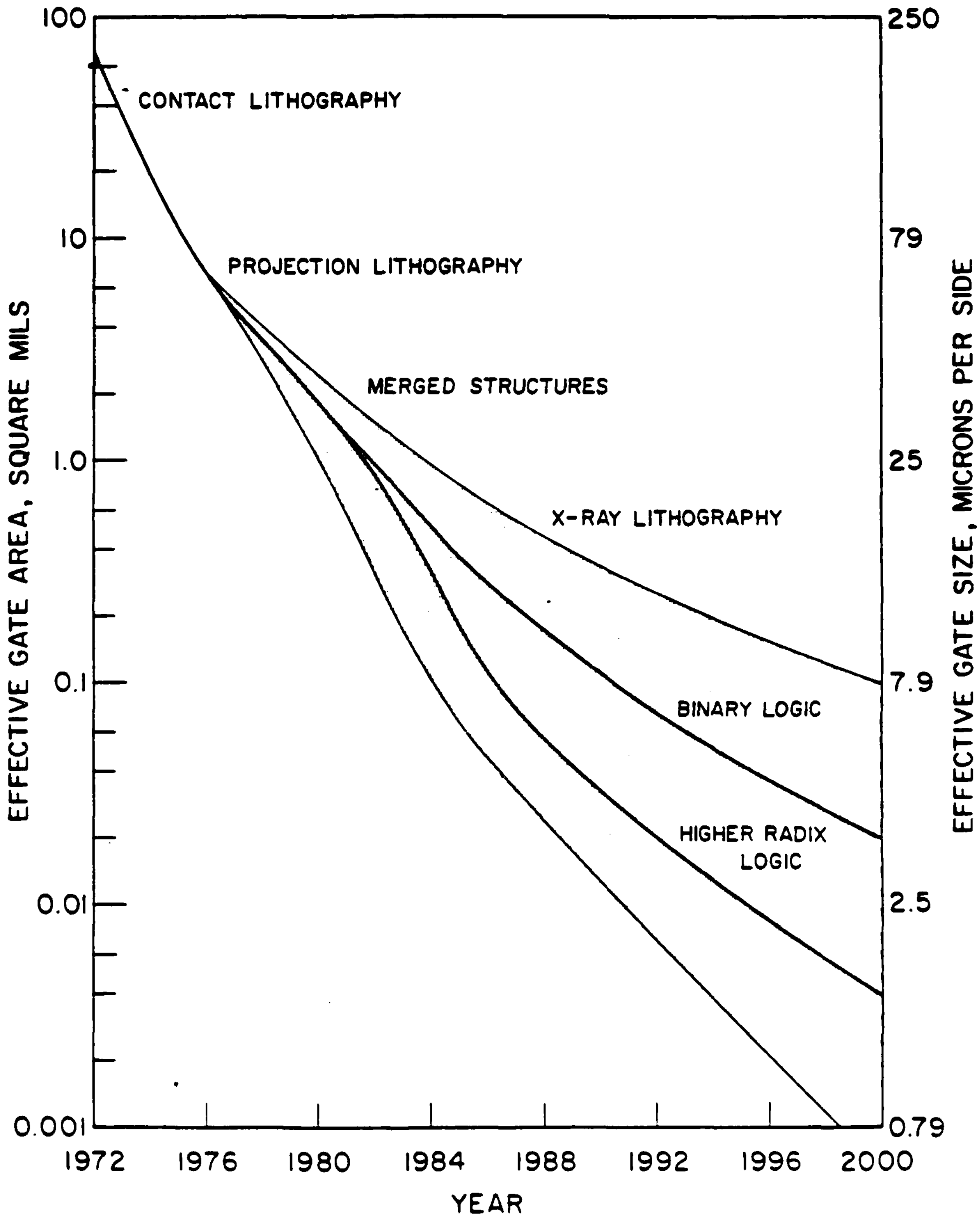


Figure 2.5

ACUMENICS

Overall, gate size appears to be evolving in good agreement with Figure 2.5 and we will continue to use this forecast as we examine telecommunications equipment in detail in coming sections.

The constant or slowly falling cost of packaged silicon combined with the dramatic decreases in gate area, results in the dramatic cost reductions on a per gate basis forecast in Figure 2.6 for future LSI systems. While cost per chip will remain relatively constant and even increase somewhat as chip size increases, the functional complexity (gates/chip) can be expected to increase substantially in accordance with Figure 2.5. This has enormous implications on the cost, size, and sophistication of future electronic systems.

2.2.2 Reliability

Present reliability levels for LSI components, including second-generation microprocessors, are about 99.95 percent per 1000 hours at a 90 percent confidence level and 70°C. As component complexity increases, new failure modes will undoubtedly arise and subsequently be corrected. Narrower line widths will require lower currents to avoid metal migration; however, device operating currents are expected to remain well below critical levels. Little change in component reliability is seen below the 99.99 percent per 1000 hour levels (90 percent confidence, 70°C) which should be achieved by the early 1980s. Extensive and special testing programs might do somewhat better but at greatly added expense. With increases in component complexity, however, these reliability levels are of great

Expected Cost Per Packaged Gate in LSI Circuits

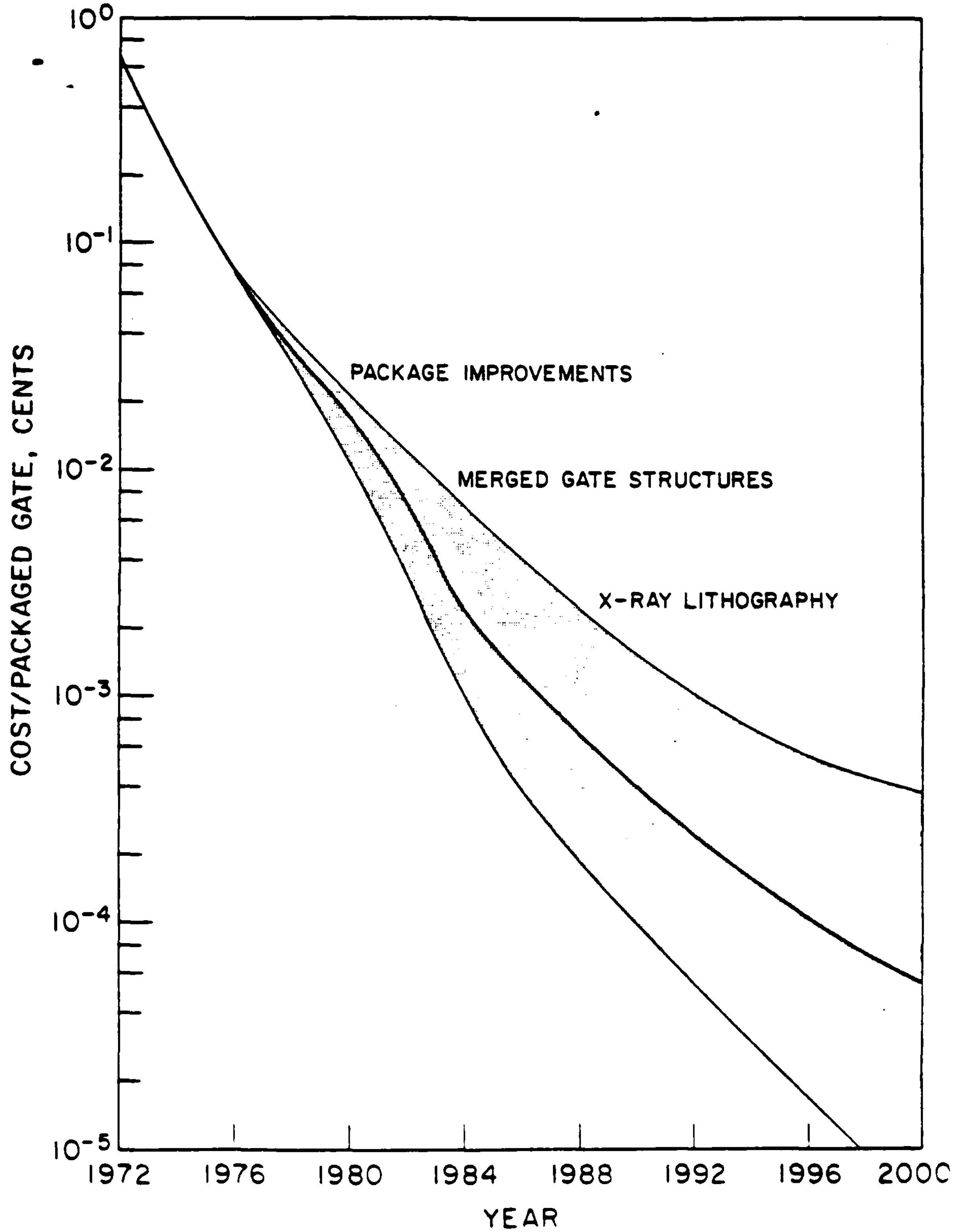


Figure 2.6

ACUMENICS

significance in future LSI circuits. Stated a different way, if we look at a relatively large minicomputer with about 10^4 gates in its processor, then in LSI technology of the mid 80s, 90 percent of such machines would still be functioning after 1000 years of operation at 70°C. Again, this level of reliability is possible during the 1980s, assuming the component has been in production for several years and that extrapolations based on accelerated testing are accurate. These reliability levels are satisfactory for most applications. In situations where failures cannot be tolerated, redundancy will be used along with on-chip error detecting/correcting circuits.

2.2.3 Speed and Power in VLSI Circuits

The dramatic increase in gate density forecast in Figure 2.5 will be possible only if accompanied by a comparable decrease in the power per gate. Package limitations are expected to continue to limit the power dissipation per package to a few watts. As gate area decreases, the node capacitance in logic arrays will decrease similarly. At a constant clock rate, the power will drop on a per gate basis but remain relatively constant per unit area. In MOS, from one to two orders of magnitude in power exist before package limitations will force on-chip power gating or limit development.

Hence, to the extent that extrinsic effects limit speed, there are two orders of magnitude improvement in speed possible as sizes decrease. Decreased operating voltages should roughly compensate for larger chip areas. In I^2L an order of magnitude or less in power exists before power gating is required although some speed advantage over MOS already exists. Another order of magnitude in bipolar speed is probably possible before transit time delays slow further progress in silicon. Both MOS and bipolar technologies are expected to mirror improvements in gate area with improved speed down to levels of about 0.5 nsec/gate and then become relatively more constant with time. Power per unit area is being traded here for speed per gate. Figures 2.7 and 2.8 show the expected power per gate at constant speed and speed (delay) per gate at constant power versus time. For CMOS inverters in silicon, a minimum propagation delay of 10 psec and a minimum power-delay product of about 10^{-18} joules have been derived (20,21) as fundamental limits.

Much progress has been made recently in logic based on new materials and effects. Gallium arsenide logic gates have exhibited delays less than 100 psec and are now being proposed for some LSI applications (23,24). Josephson junction devices have shown delays less than 50 psec (25). It is unlikely that there will be sufficient motivation to widely apply these devices in LSI before 1990, although applications in high-speed interface circuits are likely during the early 1980s. During the 1990s, the use of new materials and devices is likely in extending the performance of LSI systems below 0.5 nsec.

FIGURE 2.7 POWER PER GATE AT CONSTANT SPEED IN LSI CIRCUITS

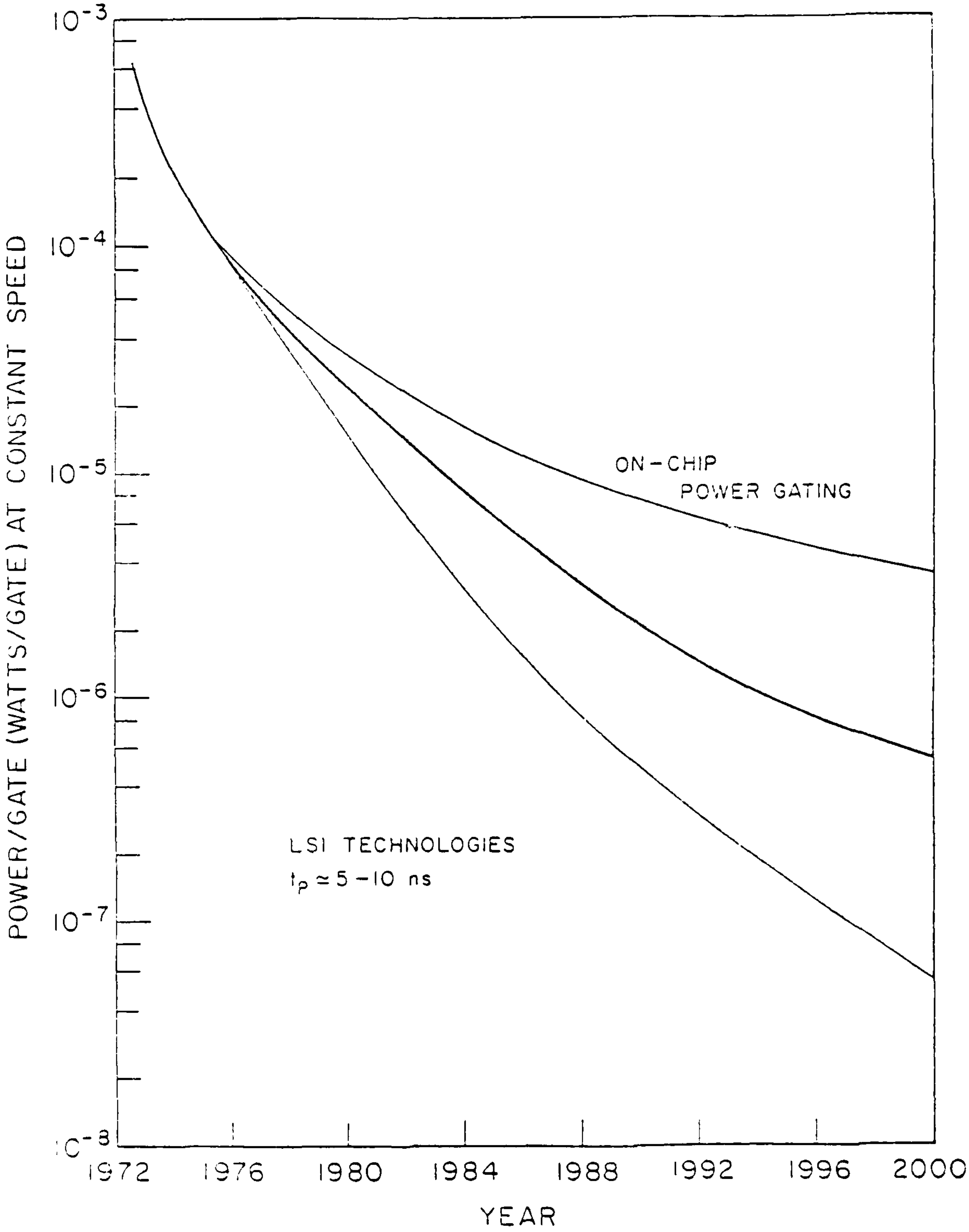
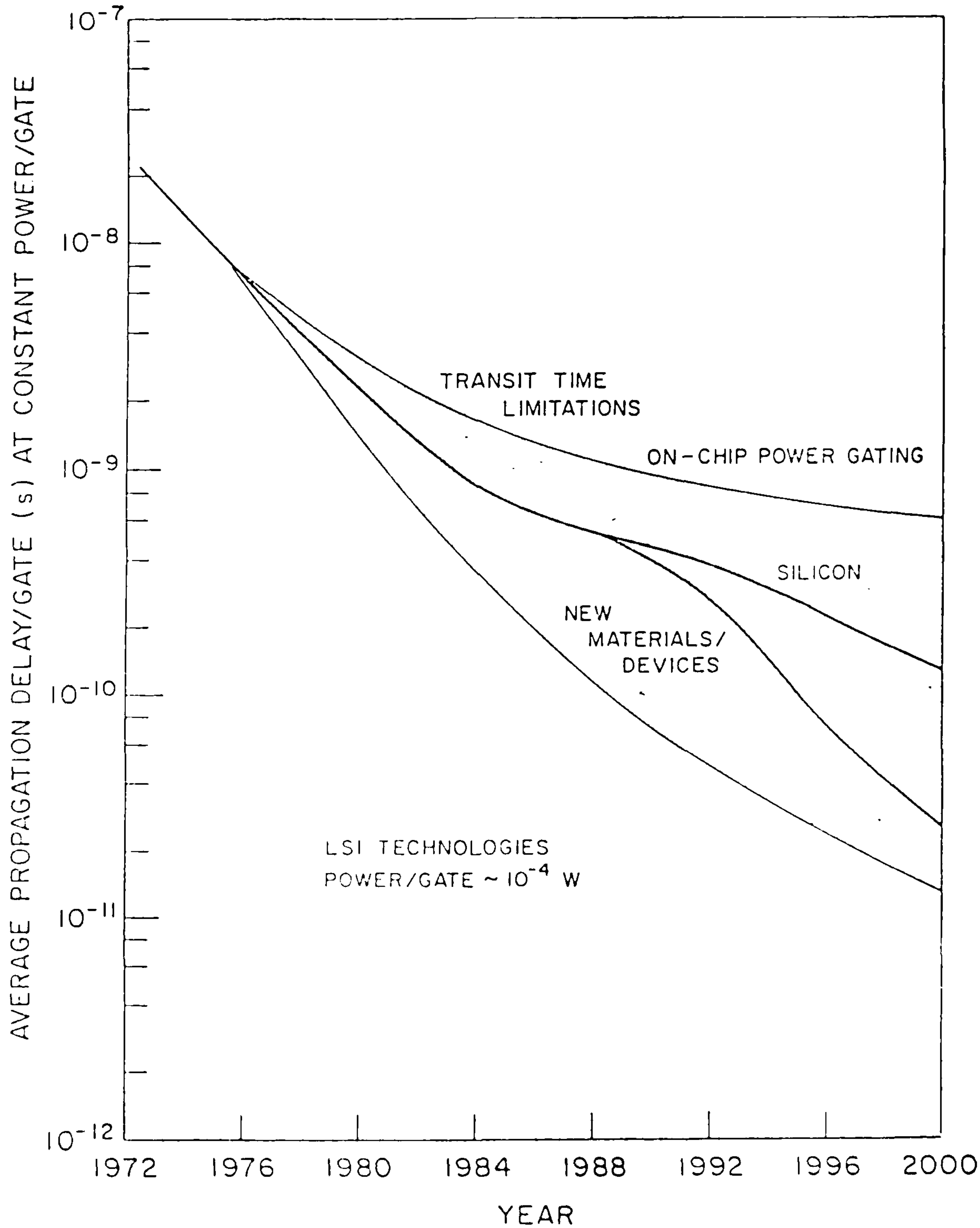


FIGURE 2.8 SPEED PER GATE AT CONSTANT POWER IN LSI CIRCUITS



The forecasts of power and speed shown in Figures 2.7 and 2.8 are consistent with the information on gate size contained in Figure 2.4, and are generally in agreement with recently reported chip data. Like gate size, these curves will be used as the basis for power and speed information as we explore telecommunications needs in future sections. It should be recognized, however, that by trading off power or density or both, substantially higher speeds are available in silicon today, with both bipolar and lightly-loaded MOS gates exhibiting propagation delays less than one nanosecond. Similarly, technologies such as CMOS offer substantially reduced power consumption. CMOS has been increasingly attractive for VLSI recently because of progress to reduce its density penalty and because of its inherent high speed, especially on sapphire substrates. Nevertheless, for the gate densities implied in Figure 2.4, the power and speed forecasts shown remain consistent and relatively accurate.

2.3 Unconstrained Microcomputer Forecasts

In this section we will examine the application of this technology to microcomputers, forecasting parameters such as cost, speed (instruction cycle), word length, and power. The distinction between microcomputers and minicomputers is expected to become increasingly difficult. In general, microcomputers will consist of one or a few LSI chips, have limited I/O, and will be applied

primarily in dedicated applications. Minicomputers will have more flexible I/O, will be general purpose in nature, but will be built around microcomputers (or possibly networks of them). In terms of performance, microcomputers are already replacing older minicomputers, and minicomputers will soon replace many present large machines.

The block diagram of a typical microprocessor is shown in Figure 2.9. An instruction cycle is composed of several clock periods during which the processor transmits the address of the next instruction from the program counter to memory, receives the instruction from memory, decodes it, and finally executes the instruction using the Arithmetic Logic Unit (ALU) to perform computations on the binary data. Nearly 200 different microprocessors are now available worldwide. Most recognize 50 to 100 different instructions and contain several thousand transistors. Figure 2.10 shows a third-generation microprocessor, the eight-bit Intel 8085.

2.3.1 Microcomputer Cost

Most first- and second-generation microprocessors were in the 30 K to 40 K square mil range in chip area. Third-generation designs, containing significant amounts of memory on-chip (e.g., microcomputers) became available in 1977 with chip sizes in the range of 50 K to 60 K square mils. Given at least three years in production so that mask-limited yields are approached and by 1980 such microcomputers could exhibit a high volume price of less than four dollars. By 1985, with an improved plastic package, the same

Block Diagram of a Typical Microprocessor

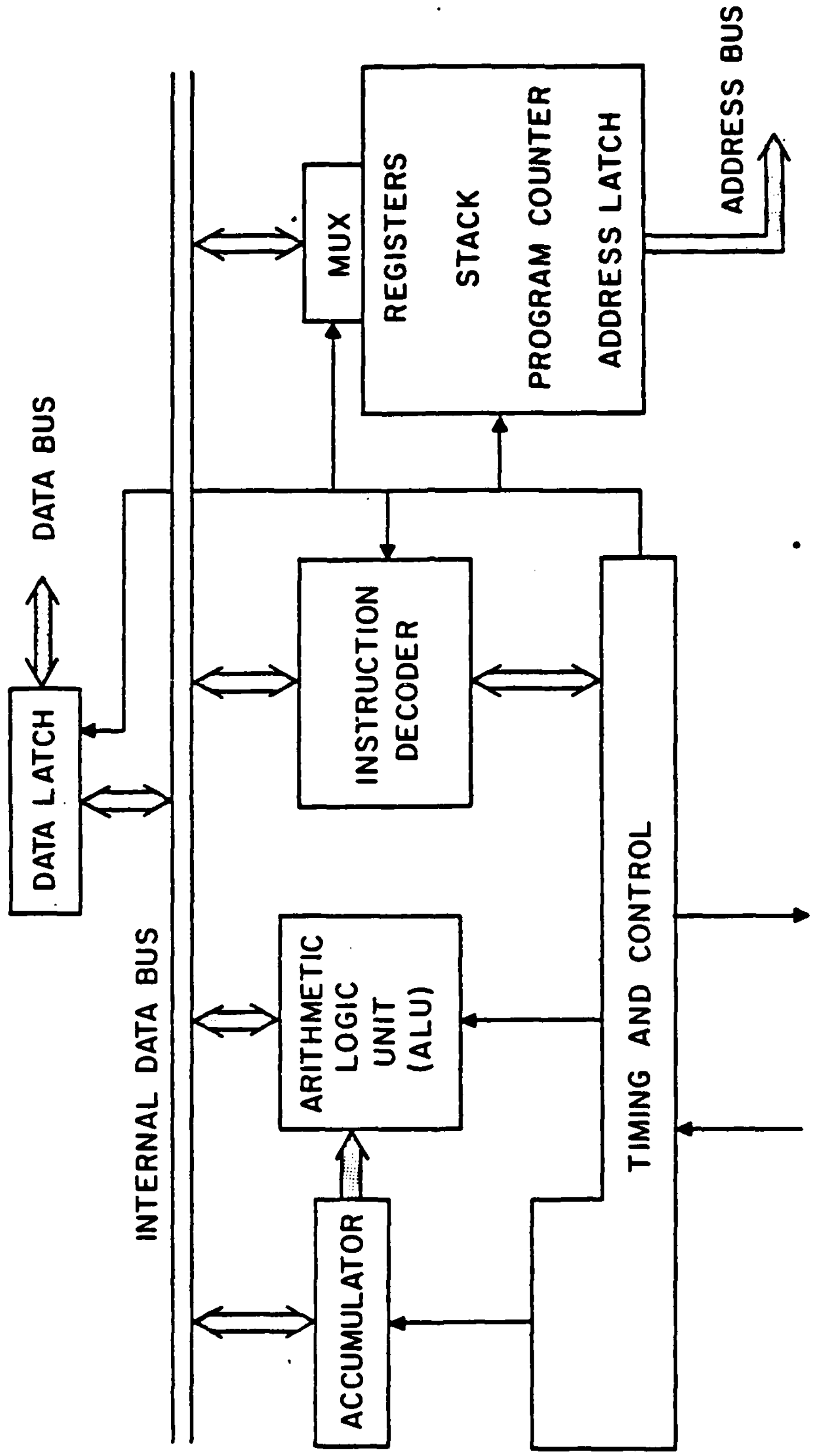
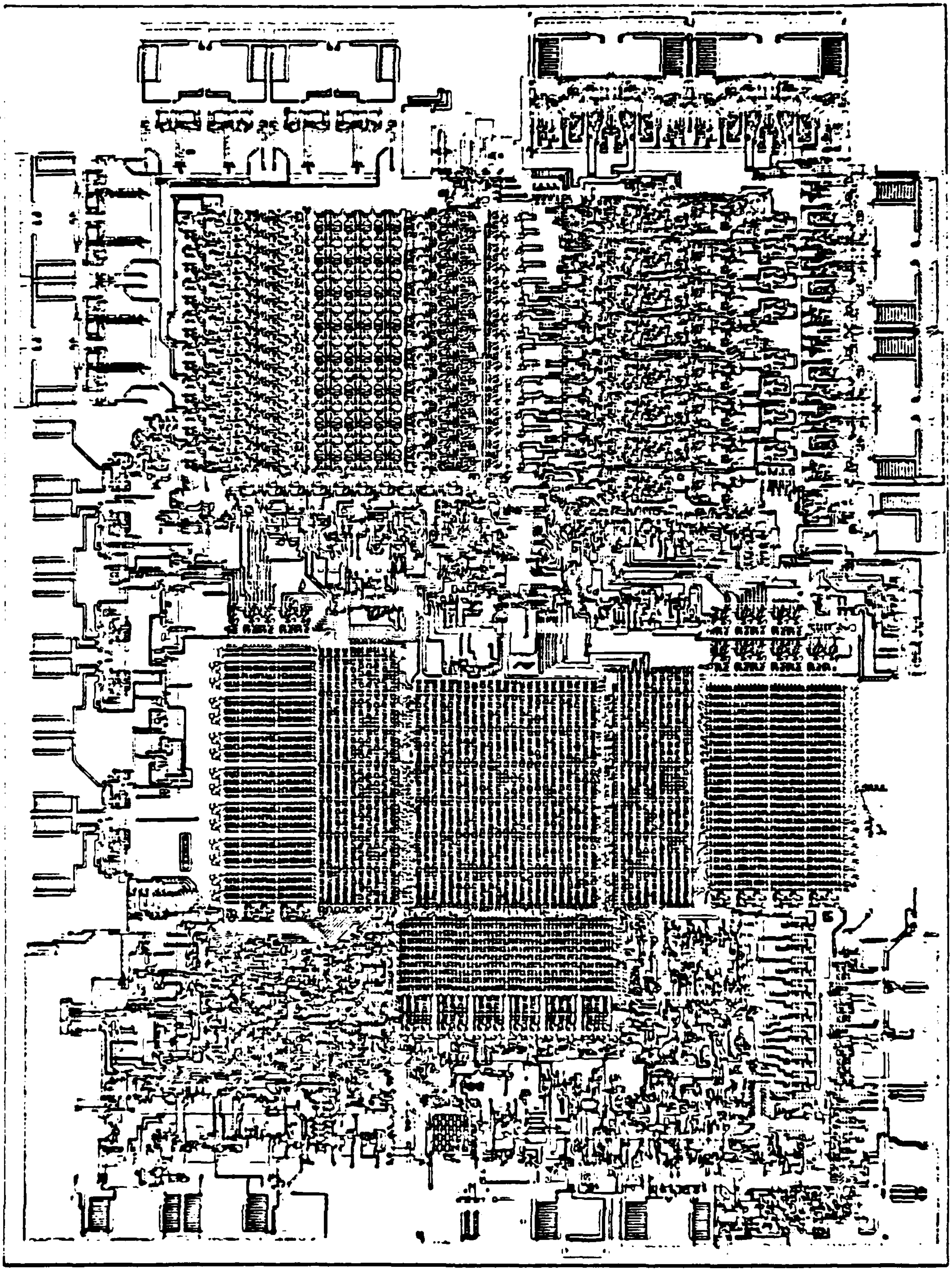


Figure 2.9

Is an example of 1977 microprocessor technology. Fabricated in nMOS, the die size is 16x222 mils.



ACUMENICS

Figure 2.10

processor in high volume (excluding marketing and distribution costs) could be near two dollars. The low cost of these elements will have an enormous impact on many areas of telecommunications, particularly in the terminal and switching areas. Although the packaged system cost may continue to exceed the LSI chip cost by one to two orders of magnitude, the system cost will nonetheless drop significantly by virtue of reduced chassis size, power, and interconnections, and it is the LSI chip that will make the system practical.

As chip sophistication increases, the design problem becomes formidable, and chip design is rapidly overshadowing chip fabrication as the major challenge to VLSI. The design problem is discussed further in Section 2.6 along with some software considerations. In telecommunications, the large application volumes can fortunately support major chip developments, although design costs may take longer to recover in the future.

2.3.2 Microcomputer Speed (Instruction Cycle)

The instruction cycle (register-register add time) of a microprocessor is typically composed of several clock periods, each of which is several logic delays long. The resulting cycle time is typically 20 to 40 times the individual gate delay for a serial processor. Figure 2.11 shows the instruction cycle time forecast versus time.

When this forecast was made in 1976 there were a few bit-slice bipolar processors capable of operating with instruction cycles in the range of a few hundred nanoseconds. The MOS procesors, which were far more numerous, were in the 2-5 μsec range in cycle time. Since that time several non-bit-sliced bipolar machines have been introduced with cycle times in the 1.5 - 4.7 μsec range, and one 1600-gate bipolar processor (26) has been announced which operates with an internal gate delay of 0.9 nsec (which could place the cycle time slightly ahead of the forecast curve). Several of the new MOS processors have cycle times between 0.4 and 0.8 μsec, making them faster than many of the bipolar machines. Some of these MOS processors are also larger (16-bit) machines, and most manufacturers have opted for increased chip sophistication and longer word length in addition to higher speed. The 16-bit Intel 8086 has a cycle time of about 0.8 μsec, so that a smaller machine such as the 8-bit 8085 scaled to the same design rules and run at the same power per chip could be expected to run considerably faster than 0.8 μsec in 1980. In spite of expected continuing emphasis on a variety of design factors in addition to speed, both bipolar and MOS machines are expected to continue to improve their speed performance. In the near term, only a few bipolar machines will offer the speeds forecast in the figure. However, by the mid 80s, the MOS machines should also reach the forecast levels, and they are subsequently expected to follow them for the rest of this century.

Forecast of Instruction Cycle Time

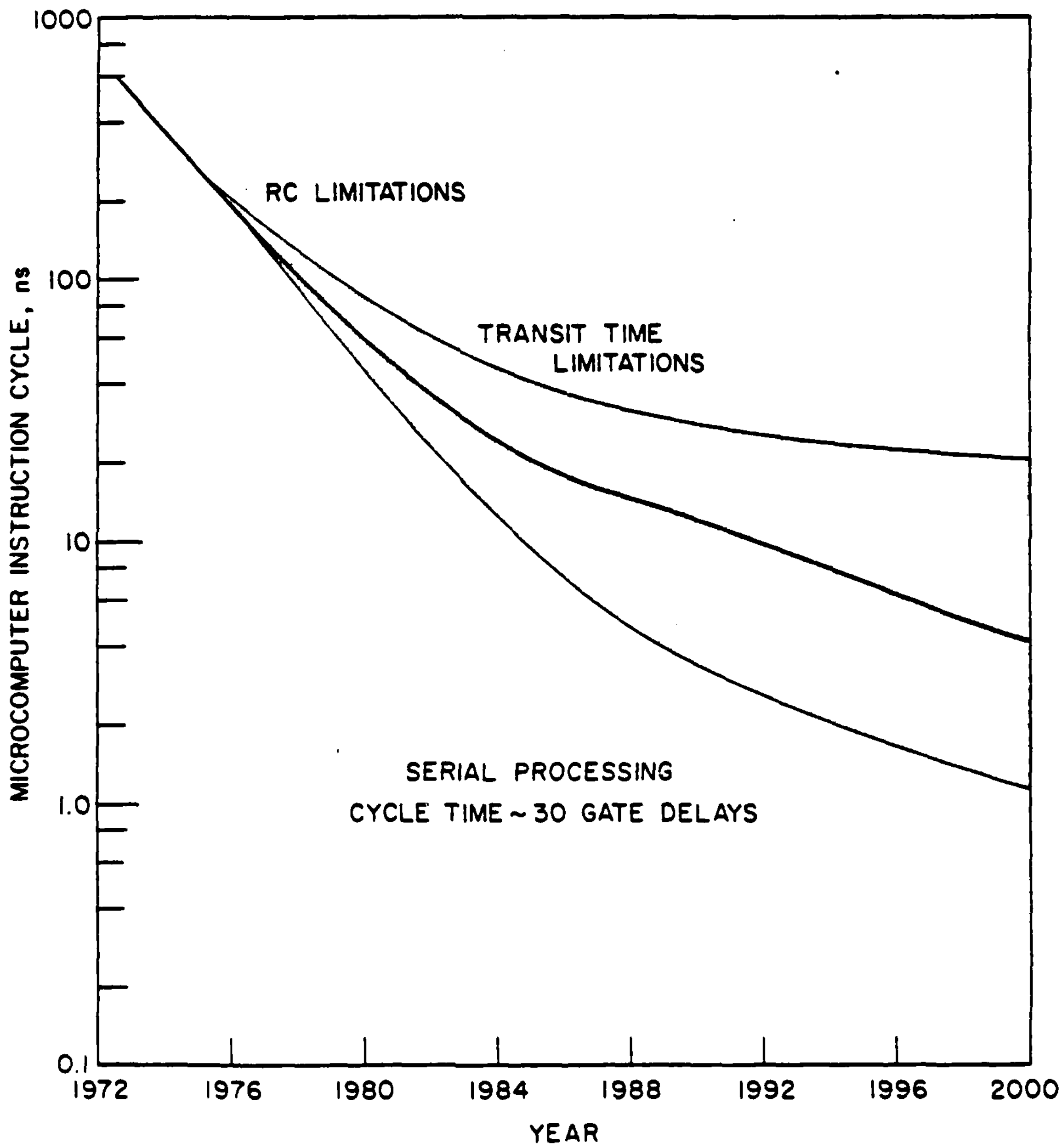


Figure 2.11

2.3.3 Microcomputer Word Length

The first microprocessor, available in 1972, was a 4-bit machine. It was quickly followed by an 8-bit machine, and in 1976, microprocessors were available with word lengths of 4, 8, 12 and 16 bits. There does not appear to be strong motivation in the industsry to develop longer word lengths, but by the early 1980s, machines having 32-bit word lengths should be available. Although little standardization in word length is discernible yet, some standardization on 8, 16, and 32 bits is likely by the early 80s.

Most of the major new microprocessors developed between 1976 and 1980 were 16-bit machines. These included entries from Intel (8086), Texas Instruments (9900), Motorola (68000), and Zilog (8000). These devices have sophisticated instruction sets including 16-bit multiplies and divides and are generally competitive with minicomputers such as the Digital PDP-11. Thus, these "high-end" microprocessors are now at the level of earlier medium-size minicomputers, with plans to go still larger in the early 1980s. At the same time, activity remains brisk in "low-end" machines for use in terminals and controllers, where 8-bits appears to be a useful standard.

2.3.4 Microcomputer Power

Microcomputer power can be expected to closely mirror power per gate for a given function at constant speed. The power for a second or third generation microcomputer having 5000 gates, each active half of the time and each biased to dissipate 0.2 mw when active is about 0.5 watts using 1976 technology. As Figure 2.12 shows, if we accept 1976 performance levels (i.e., speed), the power of a given microcomputer (redesigned each year to realize higher density) should have decreased an order of magnitude by 1984 and should fall another order of magnitude by the mid 1990s. Practically, this will not occur, since the life of any given product will be well under 10 years, and each new product generation will likely offer higher speed and greater sophistication as well as lower power. Nevertheless, the power per function will drop dramatically over the next 20 years.

2.4 Unconstrained Forecast of Memory Cost and Performance

2.4.1 Semiconductor Random Access Memory (RAM)

Semiconductor random access memory is scarcely ten years old and yet during those ten years, the number of bits available per chip has doubled virtually every year and the cost per bit has decreased accordingly. In 1978, the first 65K-bit chips were announced, and some projections forecast 265K-bit chips by the early 1980s. Semiconductor memory has provided the major focus for the evolving electronics industry and is expected to remain a major focus during the coming decade.

Forecast of Microcomputer Power

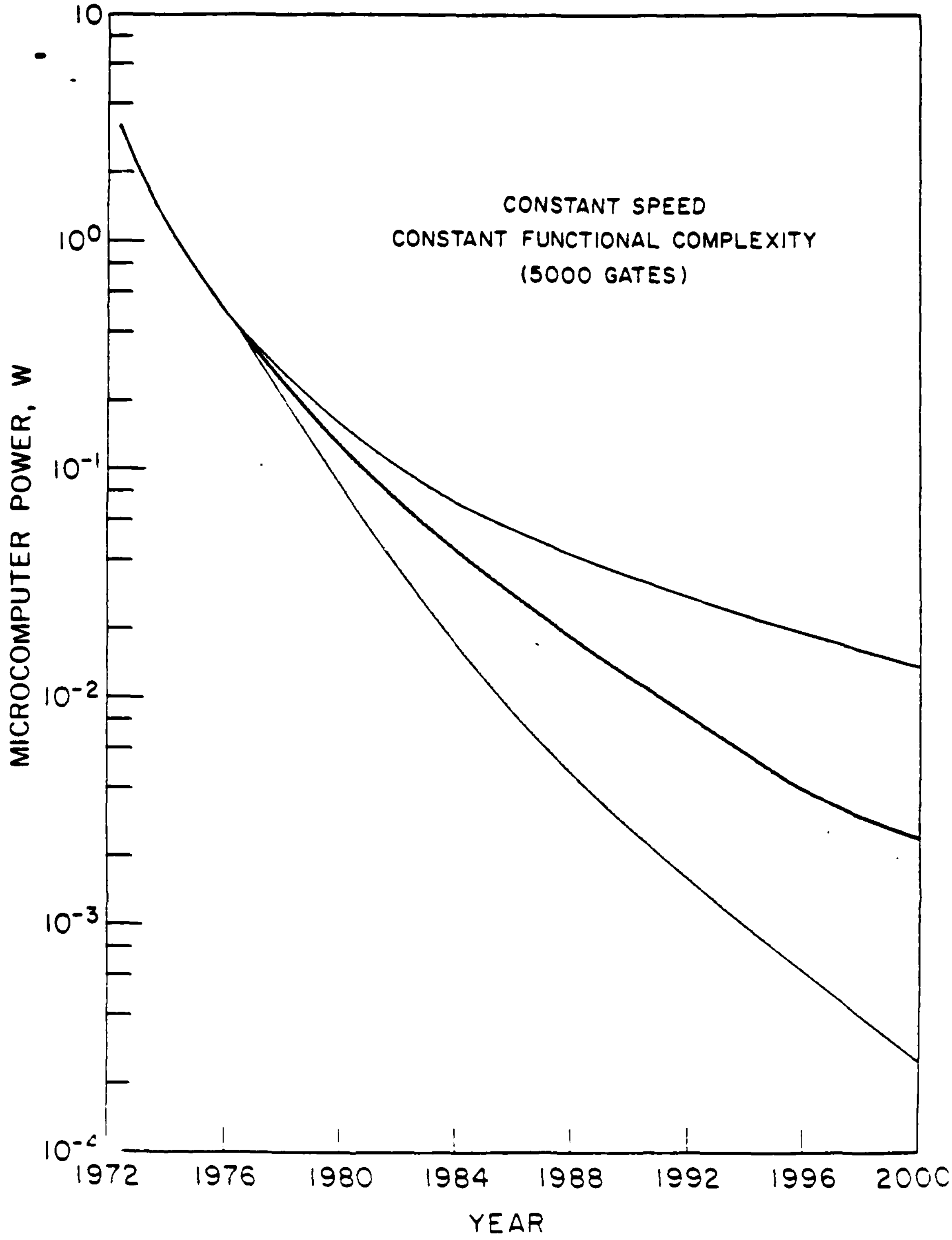


Figure 2.12

Solid-state memories come in a variety of forms, ranging from random access read-only memories (ROM), through a variety of read-write random access memories (RAM), to serial memories based on charge-coupled devices (CCDs) and magnetic bubbles. Figure 2.13 shows the evolution of these memory types during the 1970s and compares the achievable bit densities for these approaches.

The earliest semiconductor read-write memories (RAMs) were static bipolar designs using a bistable flip-flop as the basic storge cell. With the evolution of technology, denser structures became possible with some improvements in speed as well, but the basic storage circuitry remained largely unchanged. Semiconductor memory gained some acpetance for small applications, but did not penetrate the memory market sufficiently to pose any significant threat to magnetic core. About 1971, this situation changed, however, with the introduction of the dynamic MOS storage cell, in which information is stored as the presence or absence of charge on a small capacitor. The information must be refreshed (read and re-written) periodically to overcome leakage currents which slowly deplete the stored charge, and the sense amplifiers for detecting the charge must be more sophisticated. The cell is greatly simplified over the six-transistor-per-bit static cells, however, and the result is a chip with increased storage capacity. The smaller size of the dynamic cell permitted chip sizes to evolve quickly from 1K-bit (1972) to 4K-bits in 1974 and 16K-bits in 1976. Figure 2.14 illustrates this rapid progress, showing a 16K-bit dynamic RAM (1977).

Number of Bits Per Chip for Various Memory Approaches in the 1970s

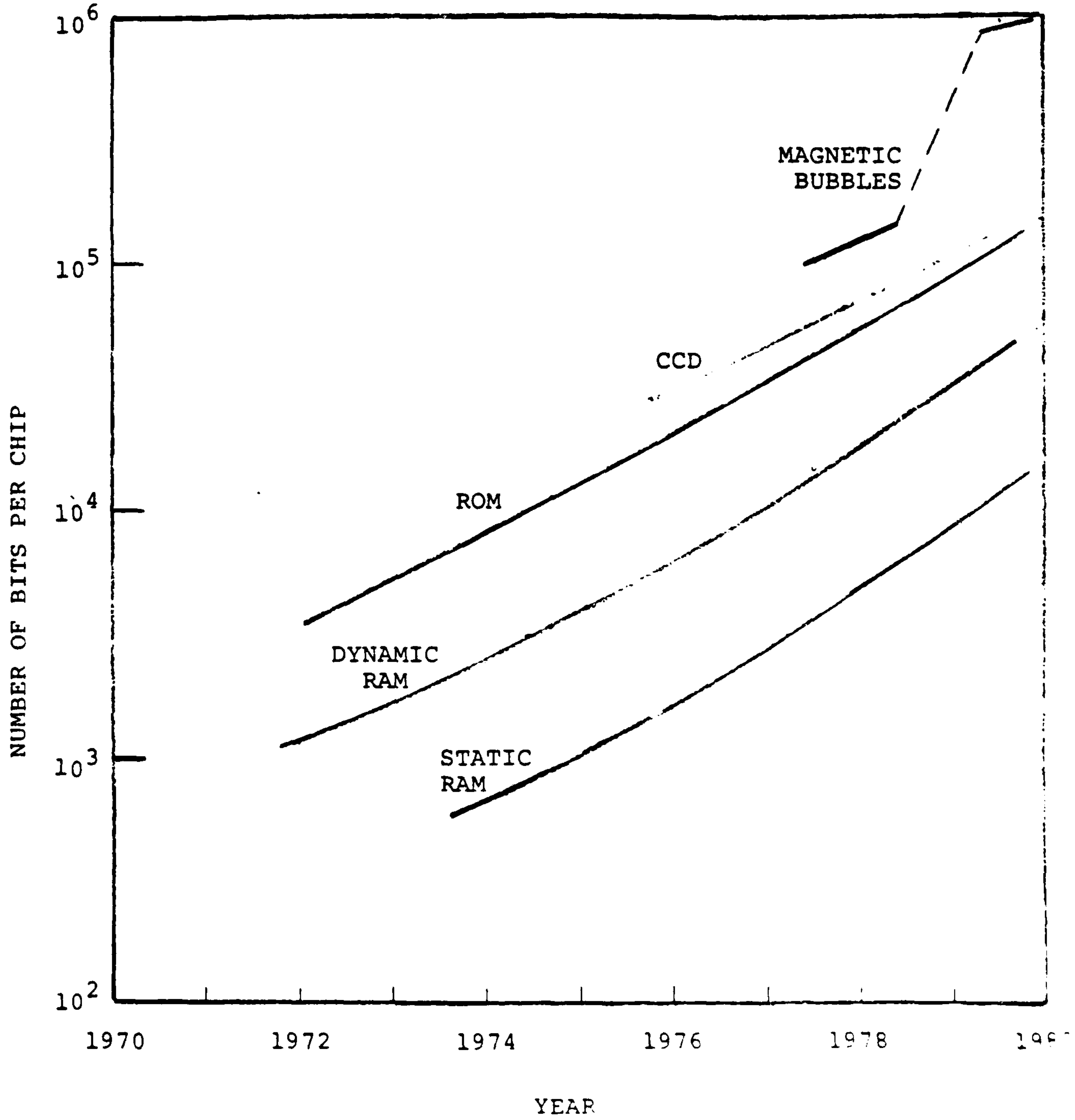
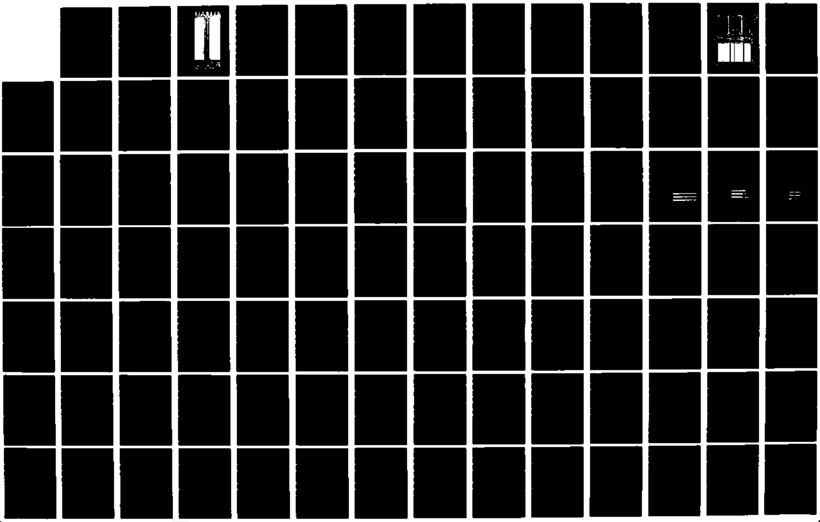
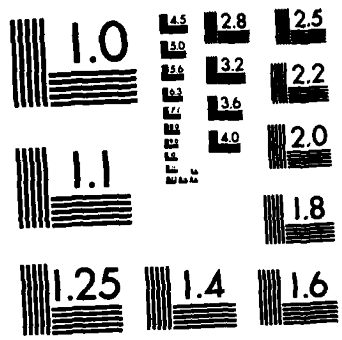


Figure 2.13

AD-A126 015 SOCIOECONOMIC IMPACT ASSESSMENT: COMMUNICATIONS INNUSTRY PHASE III TECHNO..(U) ACUMENICS RESEARCH AND TECHNOLOGY INC BETHESDA MD 02 FEB 79 FAA-APO-81-11-4 2/7

UNCLASSIFIED DOT-FA78WAI-932 F/G 17/2.1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Chips containing 64K-bits were announced in 1978 (27) and are expected to be in production in 1980. Cell circuitry remained dynamic during this period but evolved from early three-transistor-per-bit designs requiring five access lines to single-transistor-per-bit cells with only three access lines. Dynamic memories continue to be used for the highest-density designs, although there has recently been a comeback in static memories both for systems requiring very high speed and for small microprocessor-based systems. Static memories reached the 16K-bit-per-chip level in 1979 so that any density penalty compared with dynamic structures is offset by the ease with which these devices can be used in small system applications. The semiconductor memory area continues its rapid growth, employing new process technology, device structures, and circuit designs in its drive for higher density, higher speed, and lower power.

Memory Density Developments

Figure 2.15 shows the anticipated developments in bit capacity for large dynamic RAMs. The figure indicates the capacity levels expected for chips in full production; product announcements will sometimes precede these times by as much as a year or more. In 1977, the 16K-bit dynamic chips represented the state-of-the-art and were in full production. Both double-level and single-level

High Density Dynamic RAM. This 16K-bit Intel memoray (the 2117) represented the state-of-the-art in 1977. It has an access time of 150 nsec and a die size of 122x227 mils in nMOS technology.

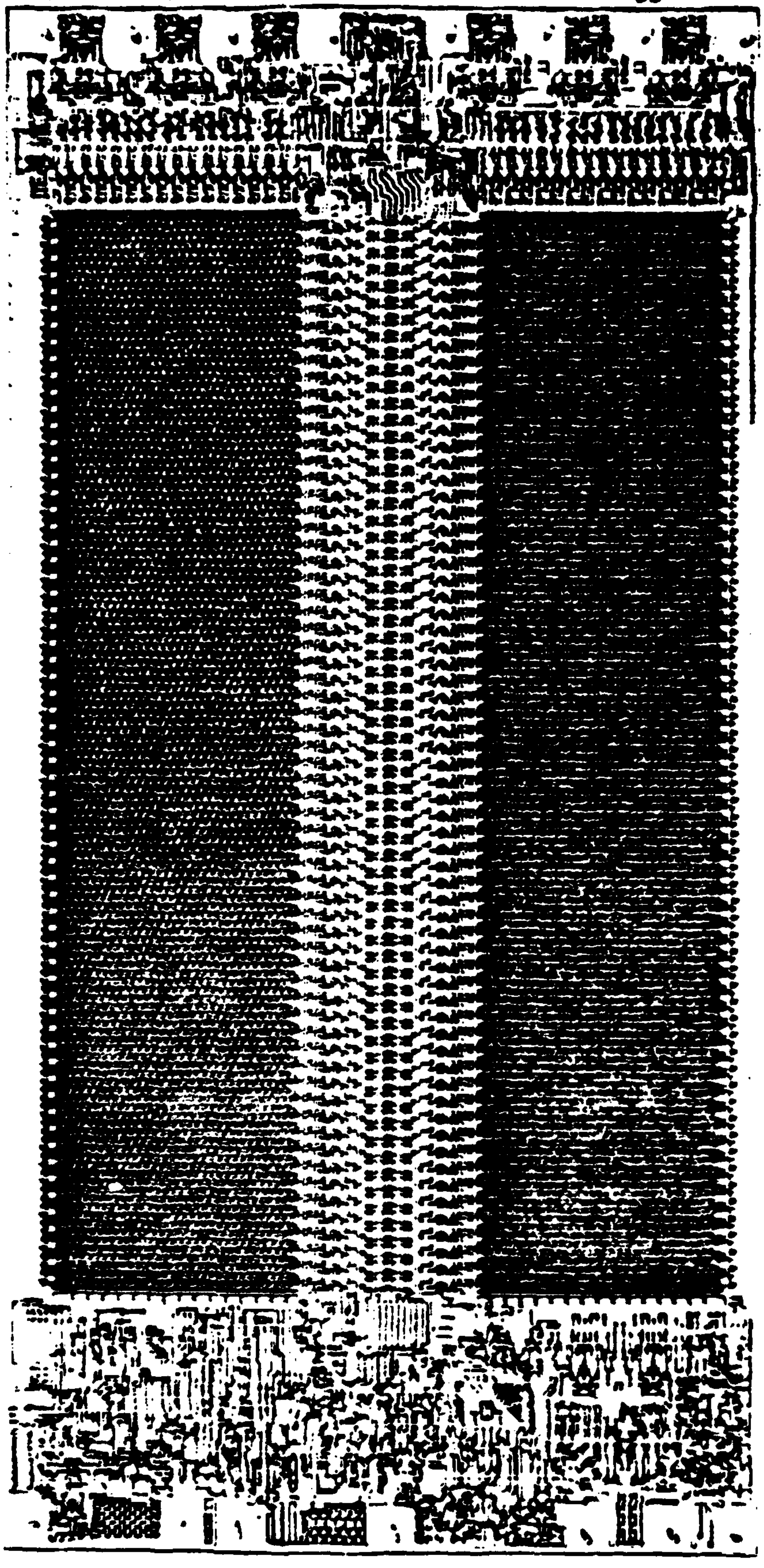


Figure 2.14

ACUMENICS

Number of Bits Per Chip for Dynamic RAM in
Full Production

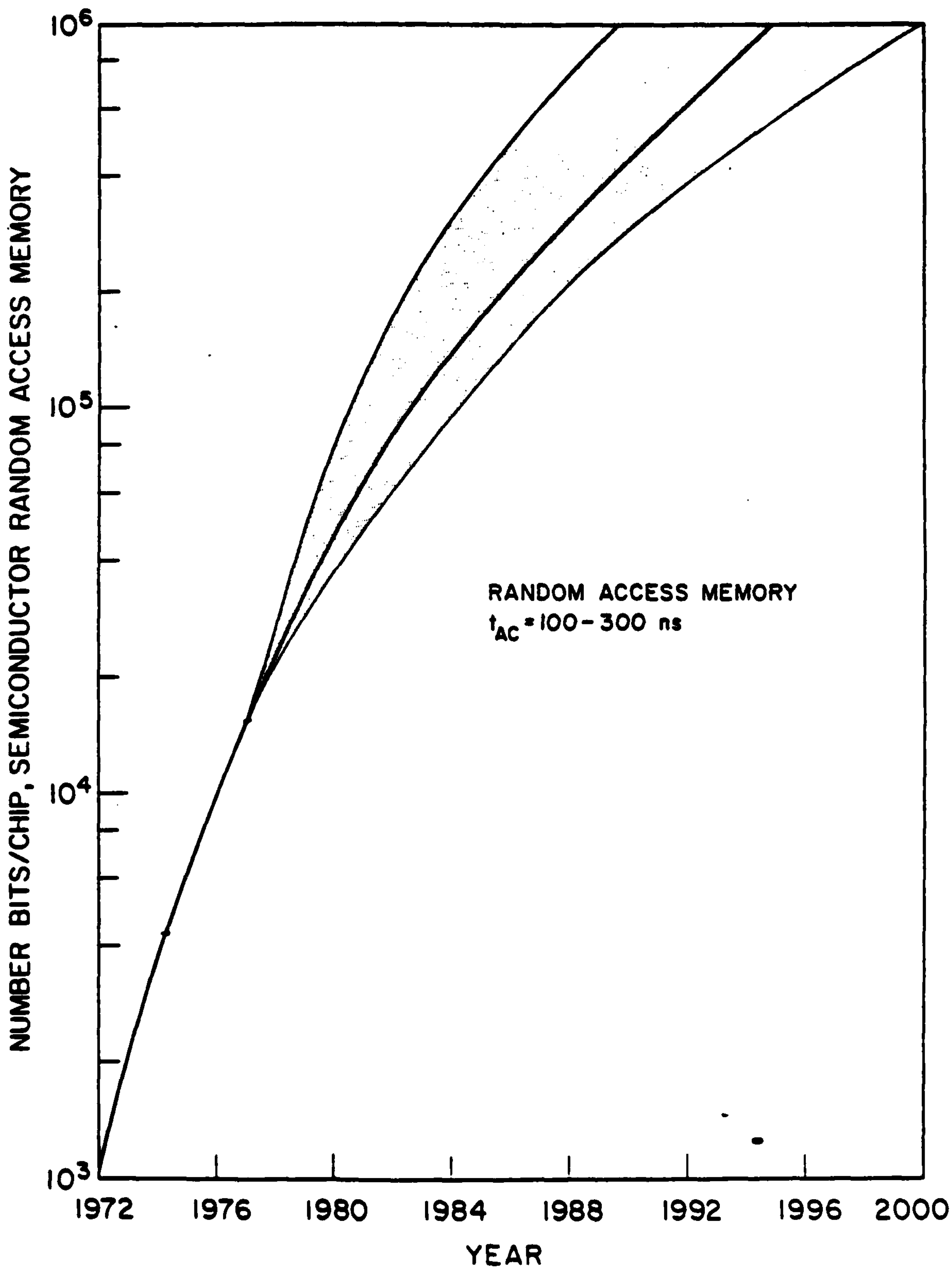


Figure 2.15

ACUMENICS

polysilicon-gate MOS processes were used in a variety of single-transistor-per-bit designs. Early chip sizes were around 180 x 200 square mils with a cell size of about 30 μm x 30 μm. As designs were iterated, significant size reductions occurred.

The initial 64K-bit chips employ a significantly smaller cell (200 to 300 μm^2), projection printed with visable or ultraviolet light and based on stacked double-poly or (in one case) triple-poly designs. Chip sizes are in the 30K-mil^2 to 50K-mil^2 (20-30 mm^2) range. The 128-bit chips, expected by 1982, should see some further cell simplifications, a larger chip size, enhanced sensitivity in the sense amplifiers, and perhaps some use of discretionary wiring for redundancy and yield enhancement. Chip size is expected to initially approach 220 x 270 mils with a cell size of 10 x 18 μm or less.

Cell density will substantially increase on the 256K-bit chips as direct electron-beam lithography is used. These chips should incorporate low voltage, low power designs with enhanced on-chip redundancy. At the 512K-bit level, x-ray lithography could reduce cell size to about 8 x 10 μm. Chip size is not expected to exceed 280 x 320 mils. At the megabit per chip level, optical addressing is certainly possible and the use of new materials and storage mechanisms is likely. Motivation for entire monolithic systems may focus efforts away from bit density as a prime design consideration.

It should be noted that most present semiconductor manufacturers are forecasting megabit RAM chips as a direct evolution of present approaches by the mid-80s. Four Japanese manufacturers recently forecast the megabit chip at dates falling between 1983 and 1986, while U.S. manufacturers are expecting such chips in the 1985-1987 time frame (28). Thus, by 1987-1989, we should expect megabit RAMs in full production. These projections are on the optimistic edge of the 1976 forecast in Figure 2.15. Present 64K RAM activity is only slightly above the forecast curve, however, and for purposes of this study we will retain this estimate, noting that it may be conservative.

Cost Per Bit of Random Access Memory

The cost per bit of random access memory has been declining rapidly since 1971 (5 cents per bit) and reached levels below 0.3 cents per bit in 1975. Further cost reductions are forecast as shown in Figure 2.16, and RAM could reach 0.01 cents per bit by 1985 or before. In the period 1971-75, cost per bit dropped by a factor of two every year. From 1976 through the early 1980s the cost per bit should decline by a factor of two about every two years and subsequently will decline more slowly. This reflects the increased bit densities, larger chips, and increased area cost of silicon. In predicting memory cost, it is assumed that the chip yield becomes defect limited approximately two years after product

Forecast of Cost Per Bit for Random Access
Memoray (RAM)

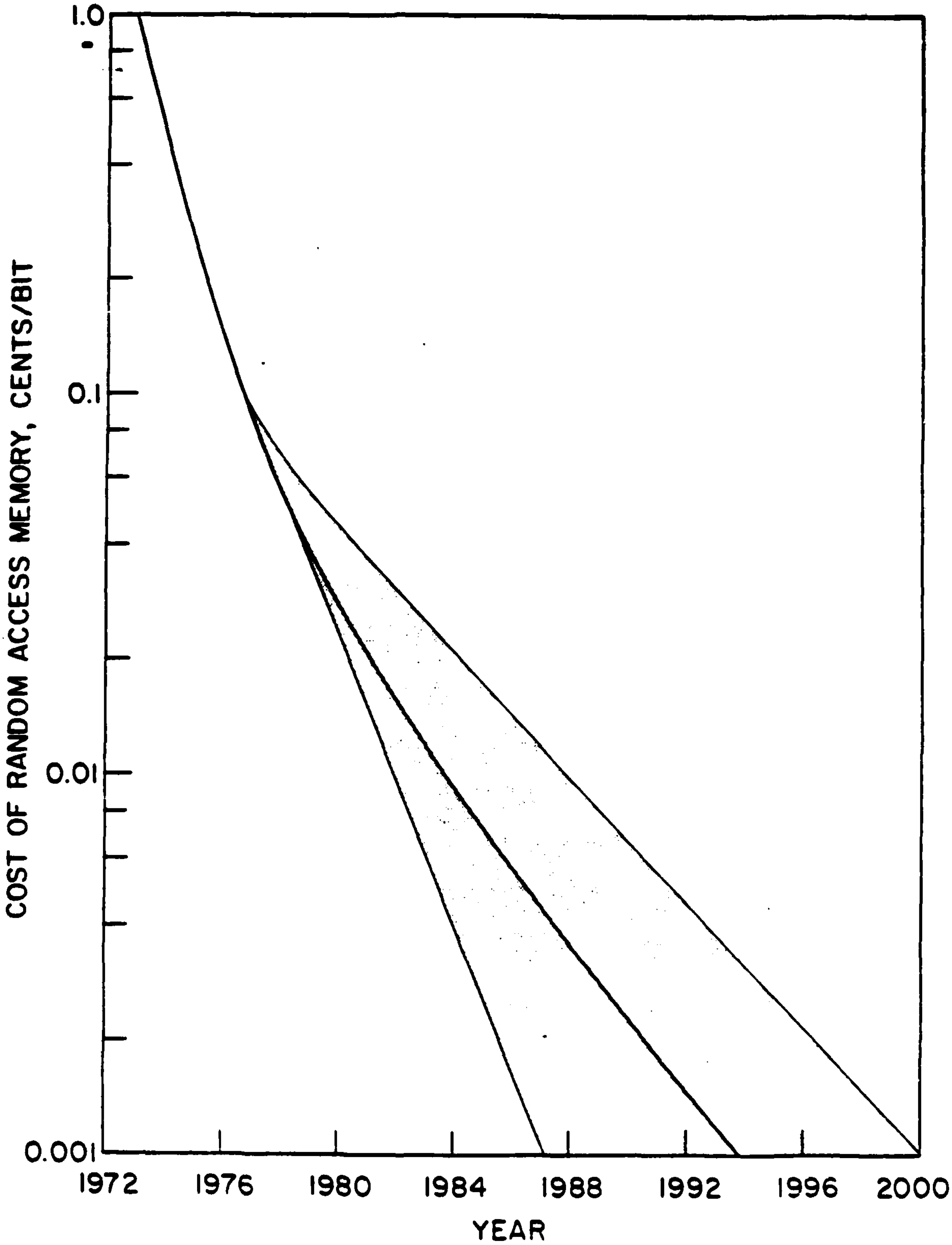


Figure 2.16

introduction, allowing time for several design iterations. The forecast assumes an evolutionary development in random access memory and that the needed developments in lithography and sense amplifier design will occur.

Random Access Memory Speed

Access time is defined as the delay between the presentation of a stable address at the memory chip and the appearance of valid data at the output. Thus, access time is related to the on-chip logic delays and especially to the capacitance of the row and column lines. For a fixed number of memory bits, the line capacitance will decrease with reduced dimensions and higher overall speed will result. On the other hand, if the number of bits is allowed to grow so that line length remains fixed, then access time will remain roughly constant unless new internal organizations are developed. New organizations are likely, and large RAMs may be organized internally as a series of smaller RAMs. Figure 2.17 forecasts the speed of a 4K-bit dynamic RAM versus time and reflects the improvements forecast in gate delays and reduced node capacitance. This forecast, originally done in 1976, still appears valid. In 1979, 16K-bit dynamic RAMs were appearing with access times below 100 nsec (29), and 4K-bit designs were still faster. Access times for 64K-bit parts were in the 100-150 nsec range, reflecting internal organizations which minimize line length and in some cases reflect

Forecast of RAM Access Time

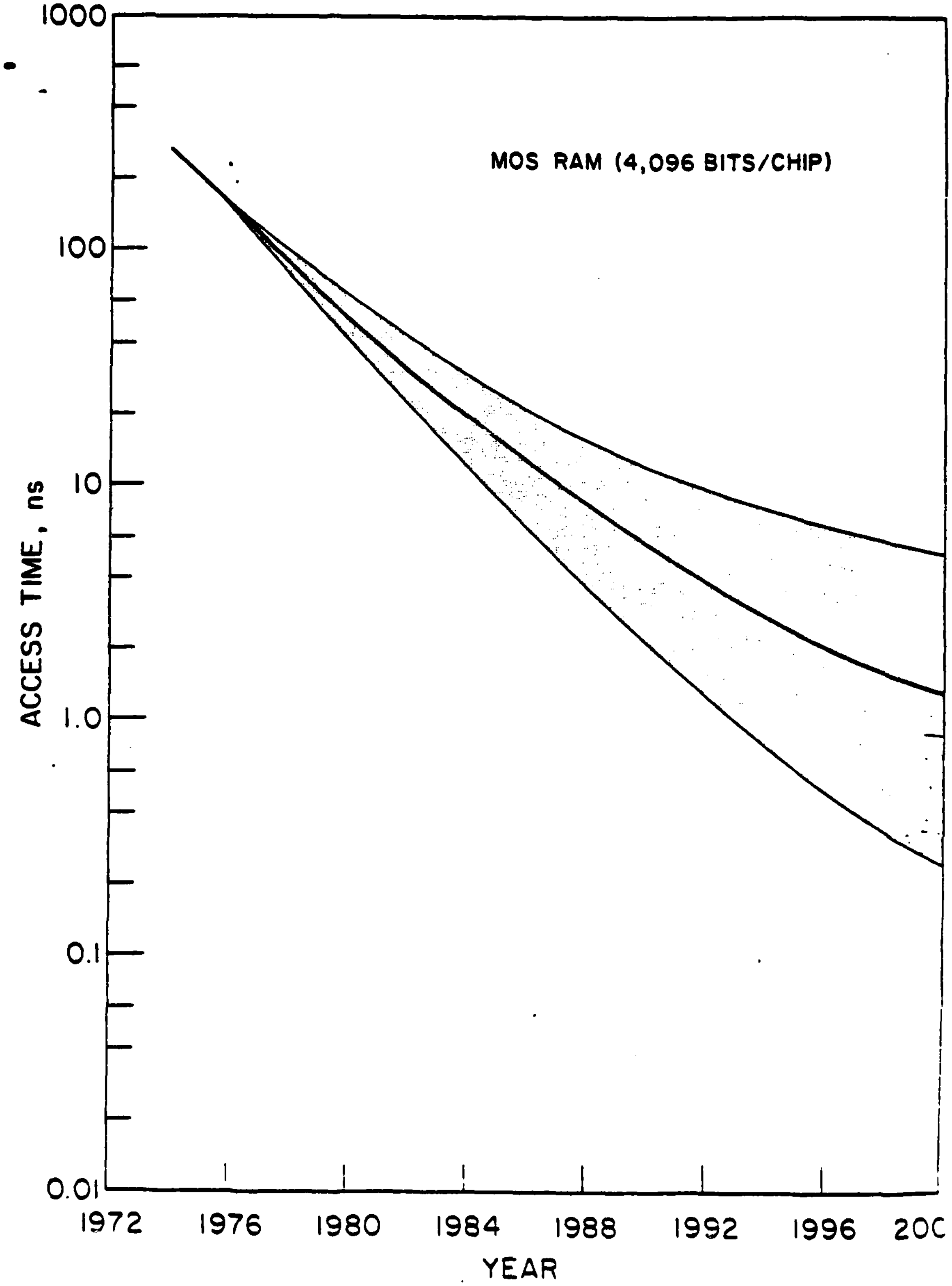


Figure 2.17

ACUMENICS

several smaller internal memory blocks. Static 1K-bit chips have dropped from 150 nsec reported for 4K organizations. This speed is comparable to that exhibited by static bipolar designs.

Random Access Memory Power

Dynamic MOS RAM as well as static CMOS RAM draws virtually no quiescent power in the storage array itself since the only DC current flow is due to leakage. In CMOS, quiescent dissipation levels of less than 10^{-8} watts/bit have been achieved. Most of the power dissipated in such memories is dynamic, i.e., associated with charging or discharging node capacitance. Since for a fixed chip size, the capacitance will remain approximately constant, the power per chip will also remain roughly constant. The power per bit will therefore decrease in accordance with cell size for a constant clock frequency. As clock frequency is increased to achieve higher speed, power per chip will increase proportional to clock frequency. The reduction of operating voltage levels, use of low-power technologies, chip partitioning, and on-chip power gating should allow the power to be held within package limits while memory speed and capacity are increased.

2.4.2 Read-Only Memories (ROMs)

Semiconductor random access memory is volatile, i.e., information is lost if the supply voltage is interrupted. While this is acceptable for many data storage applications, it is far less acceptable for program storage. Read-only memory, as its name

implies, can only be read. In its simplest form, information is written via the metal mask pattern used at manufacture. Thus, the presence or absence of a single transistor (with no storage capacitor) can be used per bit. The cell area is reduced by roughly half, compared with dynamic RAMs, while the peripheral circuitry is simplified as well. The resulting mask-encoded ROM is two-to-four-times more dense than RAM.

Mask encoding of the stored data is permanent and precludes subsequent program modifications. This can be circumvented using erasable and reprogrammable (EPROM) or electronically-alterable (EAROM) structures. The first EPROM was announced in 1971 and used a floating silicon gate to store charge (31). Erasure in these structures is by irradiation with ultraviolet light, and writing is accomplished via avalanche-induced tunneling of charge through the thin gate oxide. A single transistor acts as the storage element. Figure 2.18 shows a 16K-bit EPROM realized using this approach. Double-dielectric structures (EAROMs) are also in use for the semi-permanent nonvolatile storage of information and offer the ability to electronically erase as well as program. Their development has been slower than the pace set by EPROMs, however. Microcomputers including 8K-bits of EPROM on-chip are already a reality. Data storage for well over 100 years at 100°C has been forecast on the basis of accelerated high temperature aging.

A 16K-Bit EPROM Using a Stacked-Gate Cell (Photograph Courtesy of Intel Corp.)

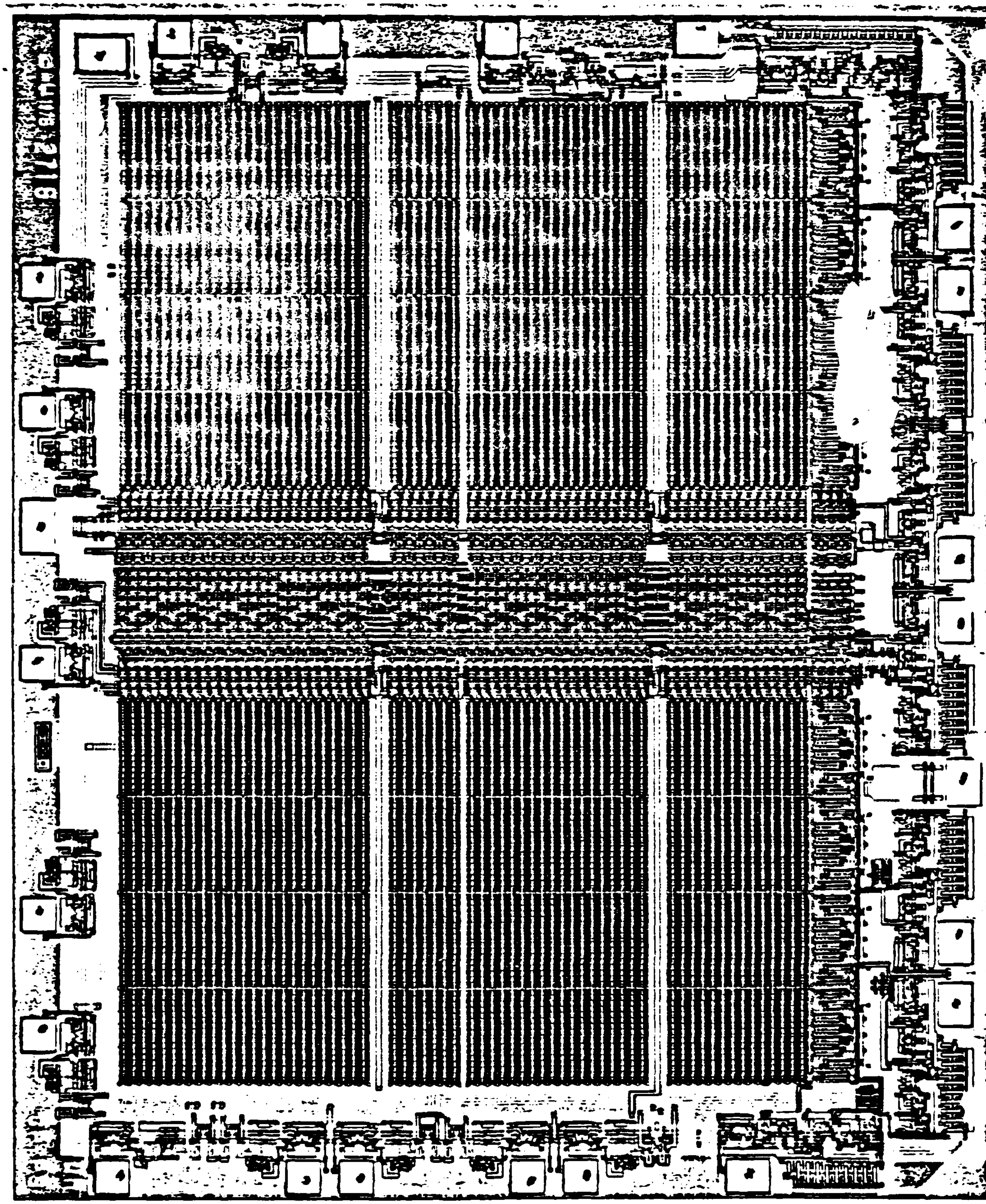


Figure 2.18

ACUMENICS

2.4.3 Bulk Storage Devices

In addition to the RAM and ROM structures already described, there are several approaches to very dense, relatively slow bulk memory which deserve mention. These monolithic structures are potential replacements for magnetic drum and disk memories, offering small size and freedom from moving parts as attractive features.

Charge-coupled devices (CCDs) (32,33) are MOS structures which store charge in the depletion layer created under the gate electrode. By arranging many gates in series, a pack of charge injected at one end of the array can be shifted along from gate to gate under the control of a multiphase clock signal. Such shift registers are very dense and relatively simple to fabricate so that they offer one practical approach to serial semiconductor memory. Normally, the chip is oranized as a number of addressable loops, and access times (typically in the microsecond range) are dependent on the length of a given loop. Since the structure is dynamic and volatile, the charge must be continuously clocked so that the information is refreshed periodically.

In 1976, serial CCD memories were significantly ahead of dynamic MOS RAMs in bit capacity; however, in the 1976-1980 period, CCD memories have done relatively little. No longer much more dense or significantly easier to fabricate than RAM, the CCD approach is being pressured by both RAM and by magnetic bubble technology and does not appear to be a major memory technology for the 1980s or beyond.

Magnetic bubble memories (32-34) utilize circular magnetic domains (bubbles) in garnet materials as the storage media and are also organized as monolithic shift registers. In development for several years, magnetic bubble memories became commercially available at the 100K-bit/chip level in 1977, exceeding CCD memories by nearly a factor of two in bit capacity. These bubble memories suffered from poor operating margins but developments continued and in 1979 megabit bubble memories were announced by Intel Corporation. This memory is organized internally as 256 4096-bit storage loops and contains an extra 64 loops used as spares to allow acceptable megabit yields to be achieved. The memory, like all bubble memories, is non-volatile and exhibits a 200 KHz maximum data rate and an average access time of 20 nsec. This chip should soon be joined by others from Texas Instruments and Rockwell.

The relatively sudden rise of the bubble memories to the megabit level was generally unexpected and ensures that bubble memories will replace disk memories for small and medium-size data processing applications and will be a key element in providing microcomputer-based systems with non-volatile bulk storage.

Electron-beam addressed memory systems (EBAMs) replace the X-Y access scheme on the memory chip with an electron-beam addressing scheme off-chip so that the cell is considerably simplified. High density and low cost per bit result but an electron tube and its

associated circuitry is required. Applications of this technology will be primarily in very large high-speed memories of more than 10 megabits. Thus, the impact on microprocessor-based systems is expected to be minimal.

2.5 Peripheral Hardware and High Performance Analog Circuits

For large time-division electronic switching and data processing systems, the number of system parts in addition to the computer itself is not large, and the VLSI impact on system capabilities can be assessed from the microcomputer forecasts almost directly. However, for terminals and instrumentation systems the impact of microcomputers could be severely muted unless these advances are accompanied by progress in the peripheral hardware area. In this section, progress in the peripheral area will be discussed -- "peripheral" being interpreted as functionally and physically separate from the microcomputer itself. The approaches in this area are more diverse than in microcomputers and no attempt to be highly comprehensive or detailed will be made. Rather, an overview of recent and anticipated progress will be given together with references to more detailed discussions.

2.5.1 Peripheral Functions in Microcomputer Systems

The elements required in the peripheral interface area at present include a large and diverse array of components ranging from line drivers, multiplexers, level shifters, filters, and data converters which are required to change the strength, origin, and

format of analog or digital signals, to the input/output devices (sensors and actuators) required to interface between the electronic system and the physical world. From a circuit standpoint, many of these devices require a mix of both analog and digital functions, and this has tended to keep them physically as well as functionally separate from the processor.

As the capability to produce low-cost LSI circuits has improved, the number of peripheral chips required for system implementation has dropped rapidly, and there has been an increasing effort to define and develop standard interface elements for many machine-machine interfaces. The number of small and medium complexity packaged circuits required for microcomputer system implementation has decreased by roughly an order of magnitude, dropping from thirty or more in 1972 to three or four in 1977. These peripheral functions will likely become still more consolidated and standardized, and many will become an integral part of the microcomputer chip itself.

2.5.2 High Performance Analog Circuits

Efforts to develop high-performance analog circuits began in the mid-1960s. From 1965 until 1972 most efforts concentrated on operational amplifiers, which rapidly evolved through at least three product generations. The early 709 design was replaced with the 741 which has since given way to a variety of designs using FET input stages, complementary npn-pnp structures, and feedforward

techniques to improve the bandwidth available (35). Unity gain frequencies above 10 MHz are available from these monolithic designs. Many of the applications for these components have been in communications equipment, including active filters.

About 1972, the monolithic phase-locked loop emerged as a general purpose analog component and, together with the area of data converters, shifted the focus of analog activity away from operational amplifiers. Phase-locked loops (36), consisting of switch-type phase comparators, a low pass filter, and a voltage controlled oscillator, quickly found a wide variety of applications in telecommunications, including frequency division, frequency multiplication, AM and FM demodulation, and signaling (tone and frequency-shift-keyed detection). Monolithic loops have substantially improved their frequency response range, which now extends to 50 MHz. Specific applications will be mentioned in Section 3.

The area of analog-to-digital and digital-to-analog converters has recently been the focus of a great deal of work in the electronics industry and is a key element in the implementation of a variety of electronic systems. The requirements on such converters include high speed, high accuracy, and low cost, and efforts to satisfy these requirements have motivated the joining of high precision analog technology with a batch-process production environment. Since data conversion is so important to the terminal area of telecommunications, it will be described here in greater detail, while Section 3 will deal with specific applications.

A simple current-mode digital-to-analog converter (DAC) is illustrated in Figure 2.19. A signal voltage repesented as a digital n-bit word is to be converted into an analog form. Each bit of the digital representation is used to control a switch which directs current from a reference level into either the summing node of an operational amplifier or into ground. Since the summing node is at a virtual ground, the circuit avoids RC time constants and voltage settling problems and develops an output voltage

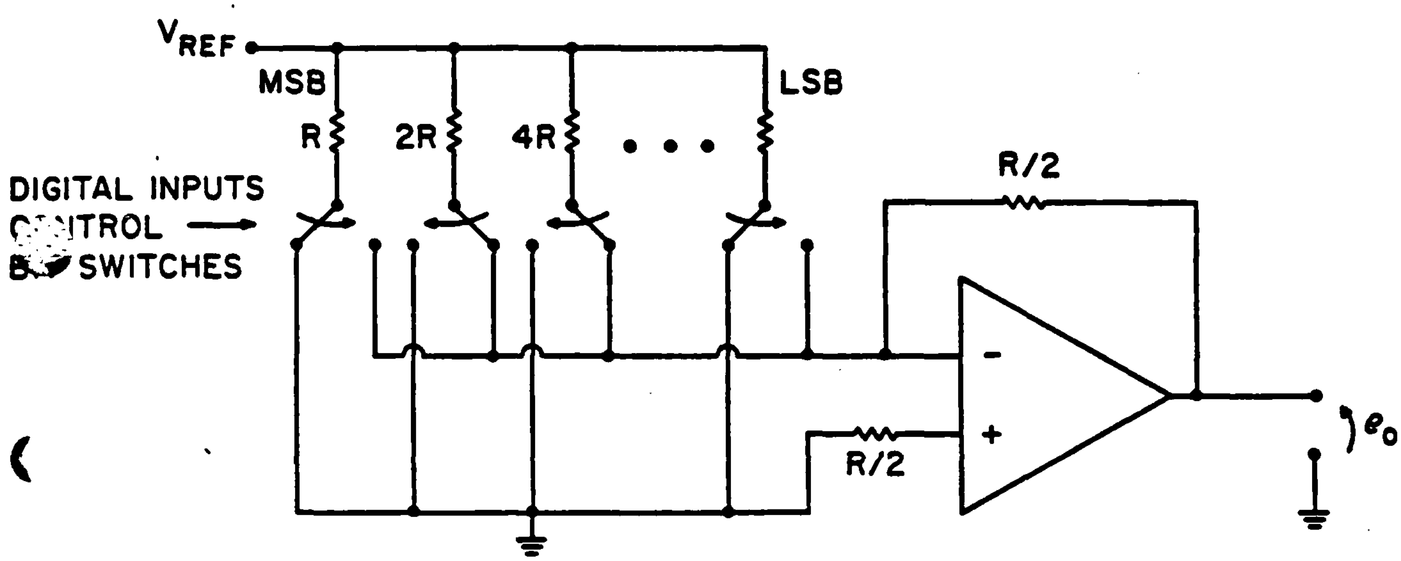
$$e_o = -V_{REF}\ (1/2\ e_1 + 1/2\ e_2 + 1/8\ e_3 + \ldots - 2^{-n} e_n)$$

where e_i, the i^{th} input bit, is either 0 or 1 depending on the input binary word. This relatively simple approach to digital-to-analog conversion has been used for most DAC applications although actual implementations are more complex than that illustrated. The output voltage can represent the desired analog signal to within one-half of the least significance bit (LSB) ($\pm$ 1/2 LSB) so that the resolution of the converter is

$$\text{Resolution} = \pm\ 1/2\ \text{LSB} = \pm\ (1/2)^{n+1}$$

For an eight-bit converter (n = 8) we have 256 possible output levels and a resolution of $\pm$ 1 part in 512. For twelve bits (the present state-of-the-monolithic-art), the resolution is 1 part in 8192 (nearly $\pm$ 0.01 percent). Achieving this resolution in a batch process has presented a formidable challenge to the industry which has been met to a considerable degree.

A Current-Mode Digital-to-Analog Converter



INPUT: 011 ...

OUTPUT: $-V_{REF}(1/4+1/8+\ldots)=e_0$

Figure 2.19

The realization of the DAC shown functionally in Figure 2.19 as a monolithic chip involves the realization of four key elements in precise and stable form: 1) the resistive ladder network, which weights the respective currents in a binary fashion, 2) the bit switches, 3) the summing amplifiers, and 4) the voltage reference. The ladder network has been a prime focus for work, since the relative scaling of the resistors determines the resolution and relative accuracy of the DAC. Both binarily-weighted and R-2R ladder networks have been used employing diffused, ion-implanted, or thin-film resistor technologies (37, 38). Thin-film Si-Cr resistors have been used for many high-accuracy DACs and have exhibited absolute and tracking temperature coefficients of -30 ppm/°C and $\pm$ 1 ppm/°C, respectively (38). Ladder networks for converters having accuracies greater than eight bits must be trimmed, and laser trimming (39) has proved a valuable production technique in the realization of high performance DACs. The selective shorting Zener diodes to allow permanent one-time adjustment of ladder accuracy is a second technique now in use (40). Using segmented organizations, resolution levels as high as 12 bits have been achieved without trims (41).

The monolithic bit switches used to steer the bit currents in the DAC must be carefully designed to maintain the accuracy of the ladder over the operating voltage and temperature range of the part. Both the binary attenuation of equal currents and the binary

weighting of bit currents are used (37) as well as combinations of the two approaches. Similarly, the summing amplifier represents an analog component whose dynamic response and linearity are important in achieving high performance from the converter.

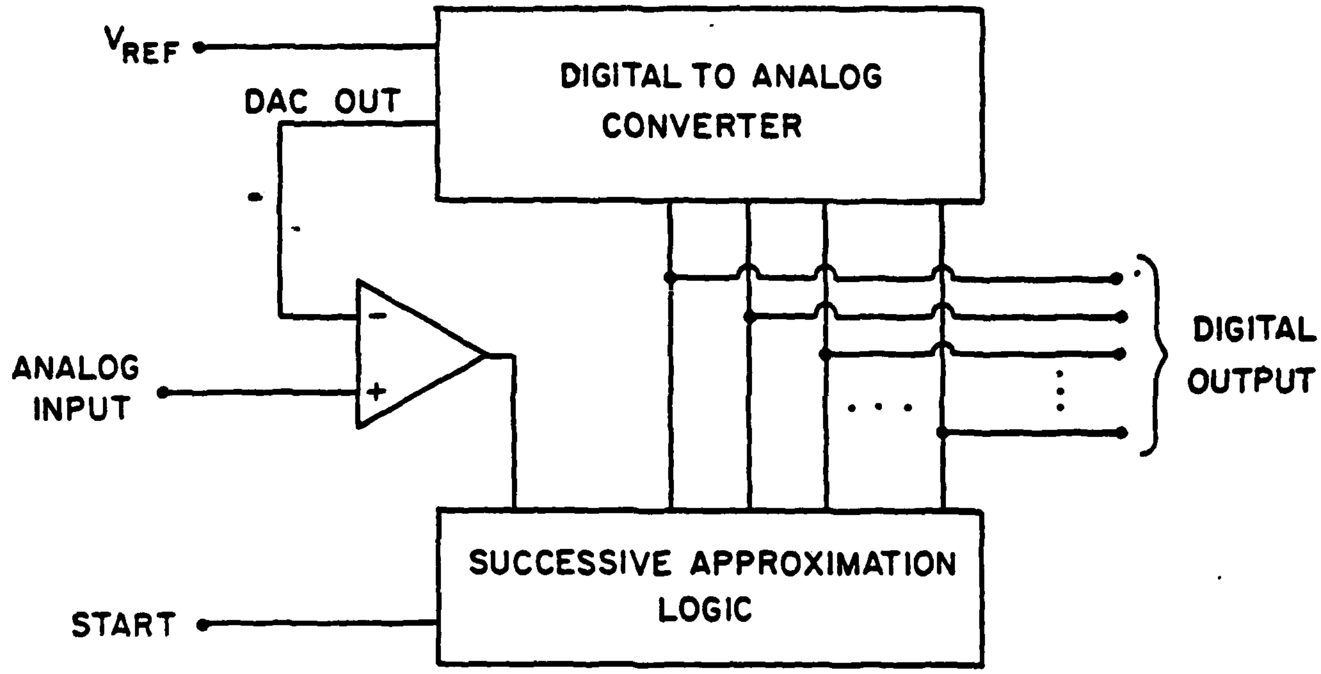
The voltage reference represents a final element whose stability is crucial to the absolute accuracy and stability of the DAC. These references in the past have been separate from the DAC chip but their integration as part of the monolithic DAC is increasingly feasible. Recent developments in band-gap (42) and buried Zener (43) voltge references, some using on-chip heaters to stabilize the chip temperature in the face of ambient temperature fluctuations, have produced references stable to better than 0.5 ppm/°C in the ambient. A variety of monolithic self-contained DACs should be available at the twleve-bit level by 1980. Comments on anticipated accuracy, speed, and cost will be given in the sections below. Integration of the DAC on the processor chip itself is now occurring (1978) in microcomputers aimed at the control area.

While most digital-to-analog converters are derived from a common resistive-ladder approach such as that shown in Figure 2.19, the approaches used for analog-to-digital conversion (ADC) have been diverse. The approaches have ranged from open-loop techniques such as analog-to-frequency, analog-to-pulse width, and simultaneous (parallel) conversion to closed-loop techniques such as ramp-and-counter, successive approximation, and dual or triple-ramp methods.

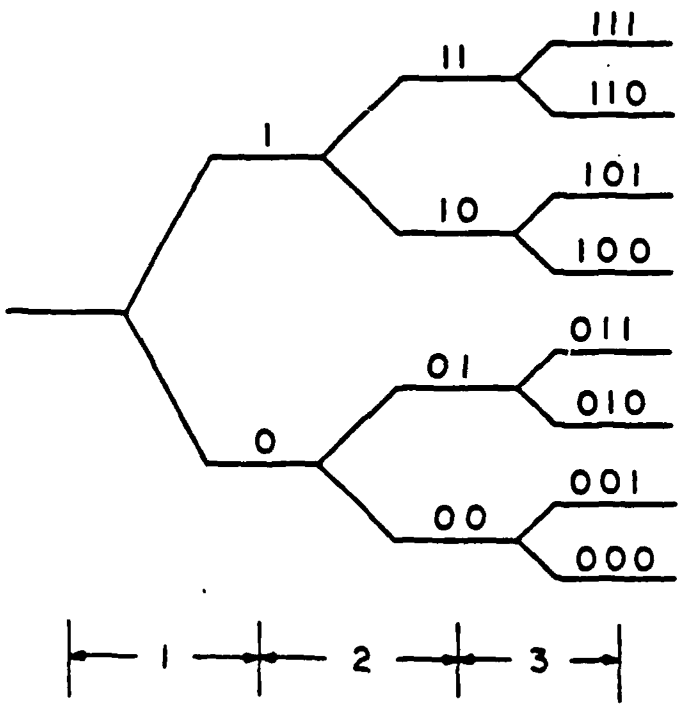
Each technique represents a slightly different compromise among speed (conversion time), accuracy (number of bits), and complexity (cost). Simultaneous converters have operated at speeds in excess of 100 MHz (44) but are limited in accuracy to less than eight bits by the required circuit complexity. They are also very expensive. Dual ramp converters require relatively modest complexity but are also relatively slow, with 2^{n+1} clock cycles required for an n-bit conversion. For a 5 MHz clock, the conversion time is more than 400 μsec for a ten-bit ADC. Such converters are widely used in digital panel meters.

The successive approximation technique represents a tradeoff among speed, accuracy, and complexity which is appropriate for many instrument applications. This technique is shown in Figure 2.20. An analog signal is presented to the converter and a START signal from the processor initiates the conversion process. The most significant bit (MSB) is set to one and the other bits are set to zero, defining a level at half or full scale. These bit settings are used as inputs to a DAC, which produces the selected analog level. This level is compared with the analog input, and if the input is greater, the MSB remains set at one; if not, the MSB is set to zero. The next-most significant bit is next set to one and the analog signal generated is again compared with the analog input, and if the input is greater, the MSB remains set at one; if not,

Analog-to-Digital Conversion Using the Successive Approximation Technique



(a)
CONVERTER BLOCK DIAGRAM



(b)
SUCCESSIVE STAGES OF CONVERSION

Figure 2.20

ACUMENICS

the MSB is set to zero. The next-most significant bit is next set to one and the analog signal generated is again compared with the analog input to determine the proper next-most significant bit setting. Successively, each remaining bit is set to one, the input is tested, and the proper bit state is determined. Thus, the successive approximation ADC requires a DAC, a comparator, and some sequencing logic. The conversion time is equal to n DAC conversion times plus n comparator delays for an n-bit ADC. In present technology this corresponds to 20-40 μsec for 10 bits.

The ADC area has been extremely active during the past five years. With the introduction of monolithic designs to replace earlier hybrid circuits, prices have fallen by nearly an order of magnitude. Fully dozens of different approaches have been tried and innovation in this area continues. Approaches using switched-capacitor arrays (45, 46) have conserved chip area and appear especially compatible with integration on the microcomputer chip itself. The diversity of approaches to analog-to-digital conversion makes forecasts of future performance difficult; however, general trends are discussed in the sections below for converters having conversion times below 40 μsec and accuracies of ten to twelve bits. Since DACs are a component part of at least one such form of ADC, they are not forecast separately.

The accuracy of monolithic data converters has improved dramatically in the past five years. Digital-to-analog converters evolved from six-bit designs, announced in 1971, to ten-bits in 1973, and in 1977 attained twelve-bits in monolithic form. Analog-to-digital converters have improved in accuracy in a similar fashion. ADCs capable of resolving at the sixteen-bit level have been reported, although these converters are not in fully monolithic form and are very slow. For successive approximation ADCs with conversion times less than 100 μs, we are now at the ten to twelve-bit level in accuracy.

Figure 2.21 shows the likely growth in converter accuracy (number of bits) for the remainder of this century. A converter offering speed comparable to today's successive approximation ADCs is assumed although beyond the very near term (1980) it is probable that some alternative technique may be used. In view of the extreme stability required as accuracy and resolution increase and the difficulty in maintaining stability over temperature, the growth in available accuracy is expected to slow considerably in the future, with fourteen-bits available in the early 80s and relatively little subsequent progress. As microprocessor capabilities improve, however, it is possible that recalibration within the end system may ease some of the stability requirements in achieving high-accuracy conversion. An additional factor slowing growth may well be the market demand for such high accuracies.

Anticipated Accuracy for Monolithic Analog-to-Digital Converters

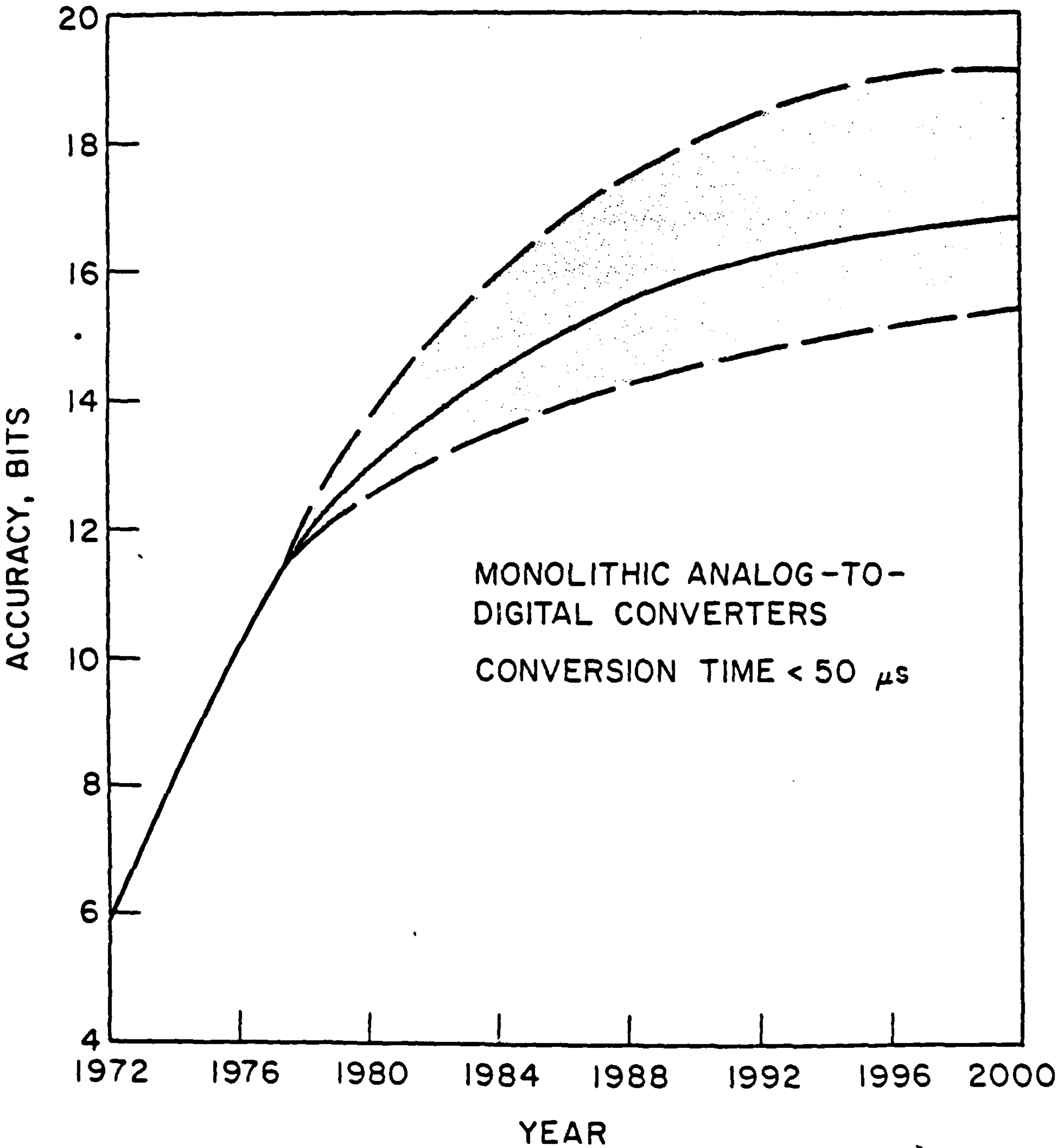


Figure 2.21

ACUMENICS

The speed (conversion time) of an ADC depends strongly on the conversion technique employed, and it is likely that a number of new techniques will be used in the future. If successive approximation is selected as representative of a present technique popular for many instrumentation and control applications, we can, however, forecast anticipated improvements in conversion time for this approach. As node capacitance decreases with reduced lithographic dimensions and, perhaps, with the application of dielectric isolation, speed improvements can be anticipated, resulting in the conversion time forecast shown in Figure 2.22. Greater improvements are anticipated in speed than in accuracy for these converters, and speed is expected to be a more sensitive function of conversion technique. For the successive approximation approach, speed is a linear function of the desired accuracy (number of bits), and at the ten-bit level, much of the present cost penalty as compared with slower techniques should soon disappear.

The cost of a converter is directly related to its yield at a given accuracy level. Figure 2.23 shows the anticipated cost of a successive-approximation ten-bit converter over time. While present ADCs of this type are relatively expensive compared with microprocessors, prices should drop below the ten dollar level during the early 1980s. In the near-term, at least, prices should change by a factor of roughly two for every two-bits of resolution added or deleted from the converter.

Conversion Time for Successive Approximation
Analog-to-Digital Converters

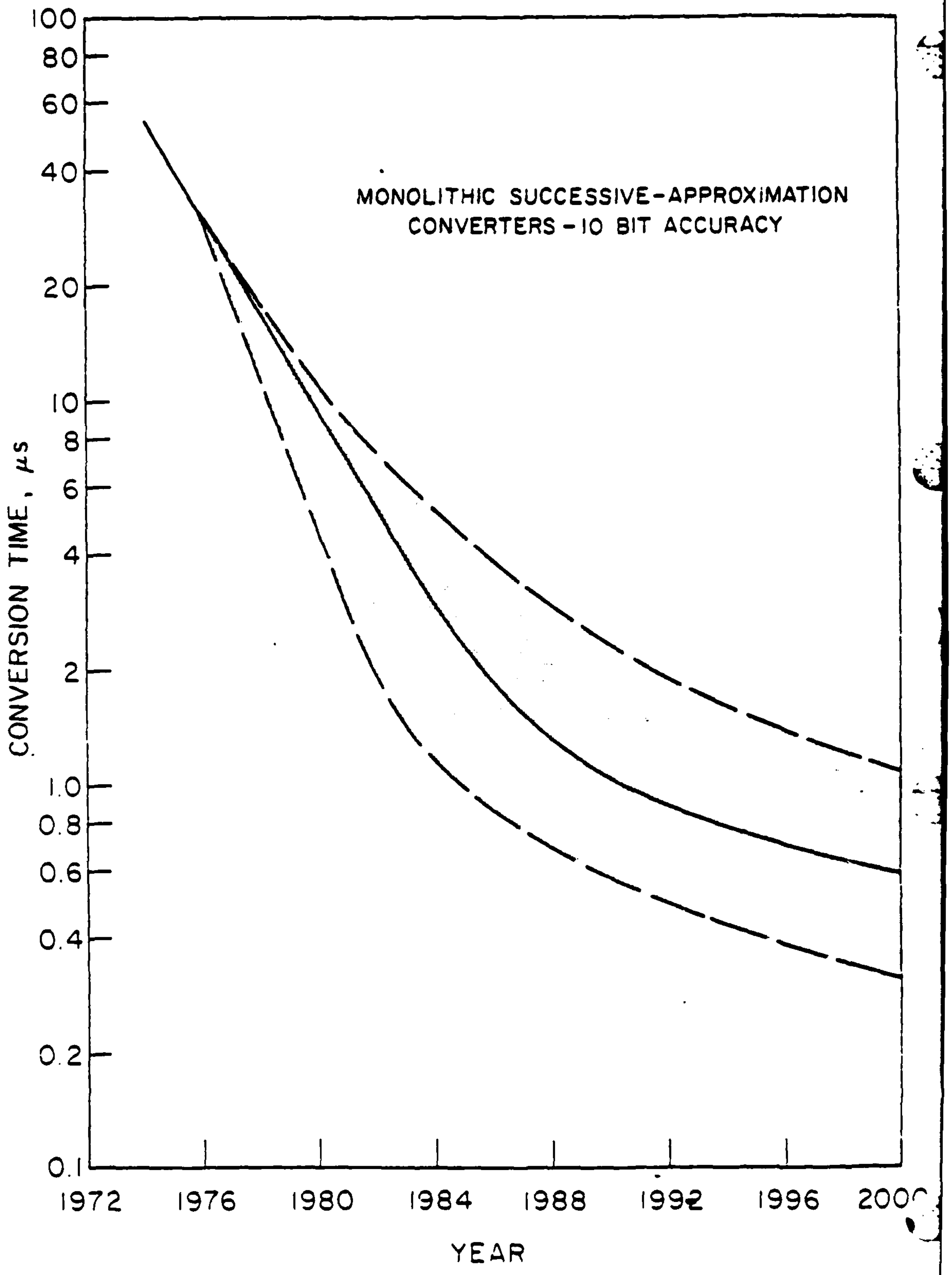


Figure 2.22

ACUMENICS

Anticipated Decline in ADC Cost as a
Function of Time

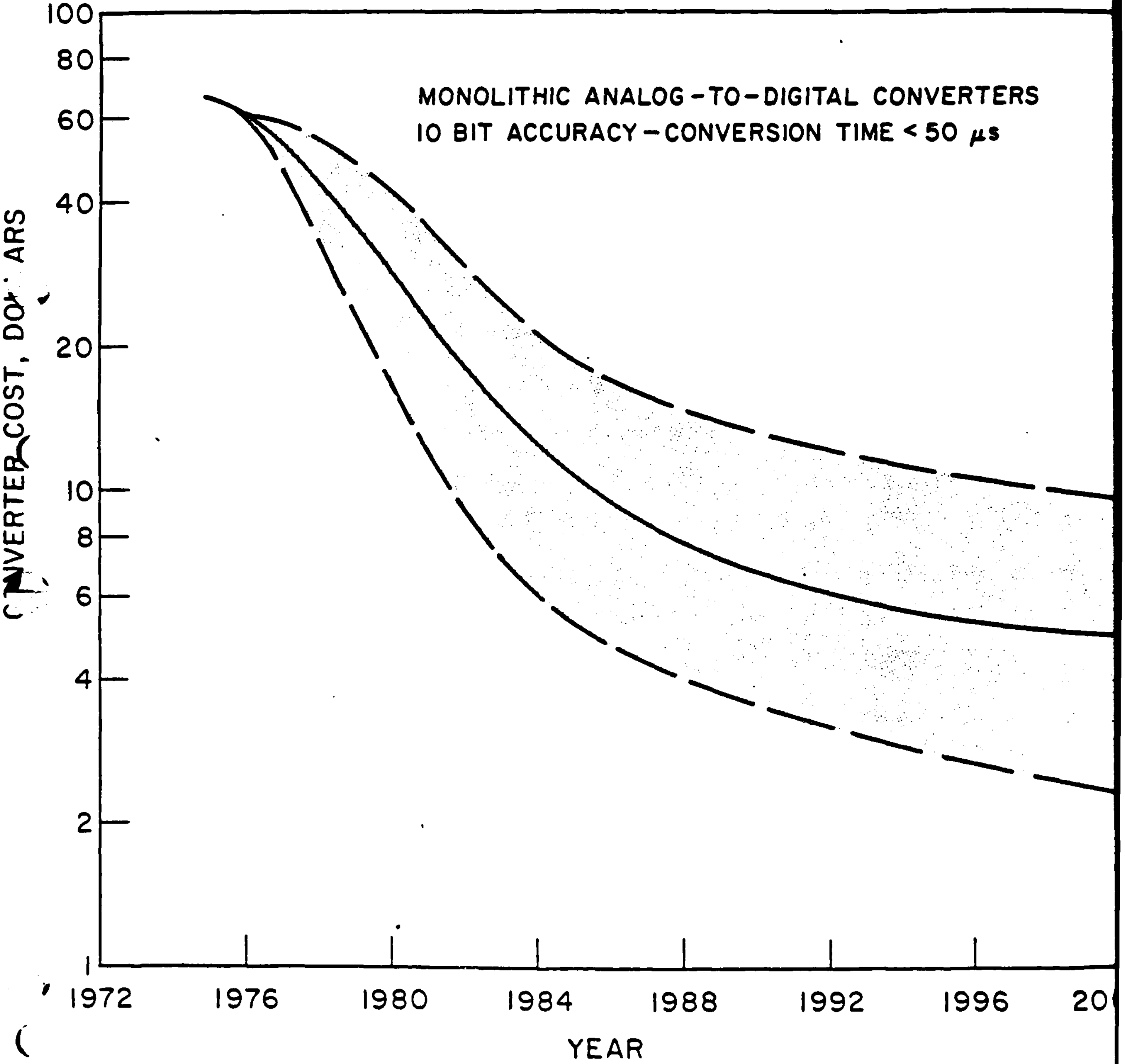


Figure 2.23

ACUMENICS

2.6 Hardware Design and Software Considerations

As microcomputer-based stored-program intelligence is incorporated into more and more systems and as the complexity of system developments grows, the significance of tradeoffs between hardware and software also increase. In systems having a fixed program and which are marketed in high volume such as input-output terminals, the portion of development cost attributed to software may be substantial, but as a component of the final system cost, software will be nearly negligible since hardware costs are repeated with each unit manufactured while software costs are not recurring. For large systems subject to frequent program modifications and sold in lower volumes, software costs may be wholly dominant in relation to hardware. For a wide range of intermediate systems, both hardware and software are important and the designer faces a spectrum of choices in trying to decide whether to implement a given function as hardware, ROM-based firmware, or software. This section will briefly discuss developments in hardware and software design which are becoming increasingly important in the creation of new systems.

Throughout the 1960s and 70s, the semiconductor industry grappled with the problem of process technology -- how to implement a given circuit in silicon. The circuit itself (and its application area and function) was typically well defined if only a means could be found to integrate it with low resulting cost and acceptably high performance. During the late 60s, the industry struggled with the problem of increasing specialization with increasing chip size.

As functions became more complex they also became more specialized, the market was reduced, and the economies of large scale integration were lost. The microcomputer solved this problem, at least temporarily, but by the late 70s the problem had surfaced again. With entire systems on one chip, the component manufacturers were forced into the systems world and the difficult problems of product definition and market development. It is now relatively clear that with the possible exception of semiconductor memory, the problem of deciding what to put on the chip has surpassed the problem of how to do it. The real challenge of VLSI may well come in the area of product definition and design and not in process development. In telecommunications, market volumes are enormous and this is one reason that major benefits from VLSI in telecommunications can be expected. As in other areas, component and systems companies must interact closely to ensure that the optimum solutions are found to the right problems. This is a real challenge in telecommunications but one which is now being met to an increasing degree.

A companion to the increasing difficulty of VLSI product definition, and the factor responsible for some of its importance, is the rapidly escalating development costs for VLSI systems. As pointed out by Moore (47, 48), in 1959 the product definition, design, and layout time required for a planar transistor was about 0.4 man-hours per month, whereas in 1978 for the 8086 microprocessor it was 184, and in 1980 for the next-generation processor it is

expected to be well over 300 man-hours per month. The definition/ design costs are doubling every two years and this, combined with a doubling every two years in chip complexity, results in a nearly constant cost per element to define, design and layout a VLSI circuit. While computer aids can help the design and layout areas, software and its development is also an increasingly important part of the design problem. The development of future VLSI chips will be an expensive proposition in which mistakes in definition, design, or layout are proportionally more critical than in the past. These areas are the real challenges of VLSI.

Expected developments in software are discussed in detail elsewhere (1). Major changes are occurring and these changes are expected to make software easier to develop, debug, understand, and alter. As memory cost decreases, the emphasis is shifting from memory efficiency to user efficiency through top-down design approaches and modularity. Structured programming, upward compatibility of software, and the use of ROM-based firmware for standard routines are expected to ease the problems of software development and decrese the cost per instruction. Use of on-chip compiliers to allow the user to interface with the machine in a task-oriented high level language instead of at the machine or assembly

language level will do much to ease the man-machine interface problem. Thus, software is expected to benefit substantially from hardware improvements, while software will provide the hardware with flexibility and the ability to perform many complex tasks with a single component.

2.7 System Implications of VLSI Technology

Past sections have examined the implications of VLSI technology for a variety of components, including microcomputers, memory, and some peripheral circuits. While great progress has been achieved in the past in these areas, still greater progress has been forecast for the future. In this section, the impact of these component developments on VLSI systems will be discussed. It is at the system level where component improvements impact the user, and it is at this level where the real significance of VLSI must be measured.

While there will be a great many different applications for VLSI technology, most of these applications can be grouped into a few broad areas. A prevous study (1) examined instrumentation and data processing applications by defining the features of a representative system of each type and then forecasting the costs, speed, power, size, and weight of that system as a function of time. This section will summarize and review the results of that study. The instrumentation system is very similar to a telecommunications terminal and the data processing system would serve equally well as the control section of an electronic switching system so these examples are not irrelevant to the present study.

2.7.1 The Impact of VLSI on Instrumentation and Control

The system considered here is an airborne instrumentation system, measuring and displaying such information as altitude, rate of climb, optimum power/fuel settings, flight diagnostics, aircraft operating parameters, and aircraft position relative to the ground. The system is illustrated in Figure 2.24, and consists of a display screen, a 262 K-bit display memory and controller, a 64 K-bit program memory (ROM), a small 4 K-bit data memory (RAM), and a 4 M-bit bulk memory.

In this system, speed must be measured in terms of the functions the system performs. Three functions examined are the times required for a display change, for data retrieval and formatting for interface with the ground, and for reading and interpreting an analog sensor. The forecast time requirements for these functions are shown in Figure 2.25 as a function of the technology-year used. The line transfer time is composed of one bulk memory access time, approximately 600 CPU cycles, one serial memory access time, and the time to transfer 512 bits to the CRT refresh memory. The microcomputer assumed for this system assumes an MOS microprocessor, making it relatively slow in near term, but also lower in power, size, and cost. For the bulk memory, a disk is used until the early 1980s when it is replaced by a bubble memory. In 1975, a line transfer was estimated at 12.1 msec and was composed of a 10 msec bulk

Block Diagram of a VLSI Instrumentation System

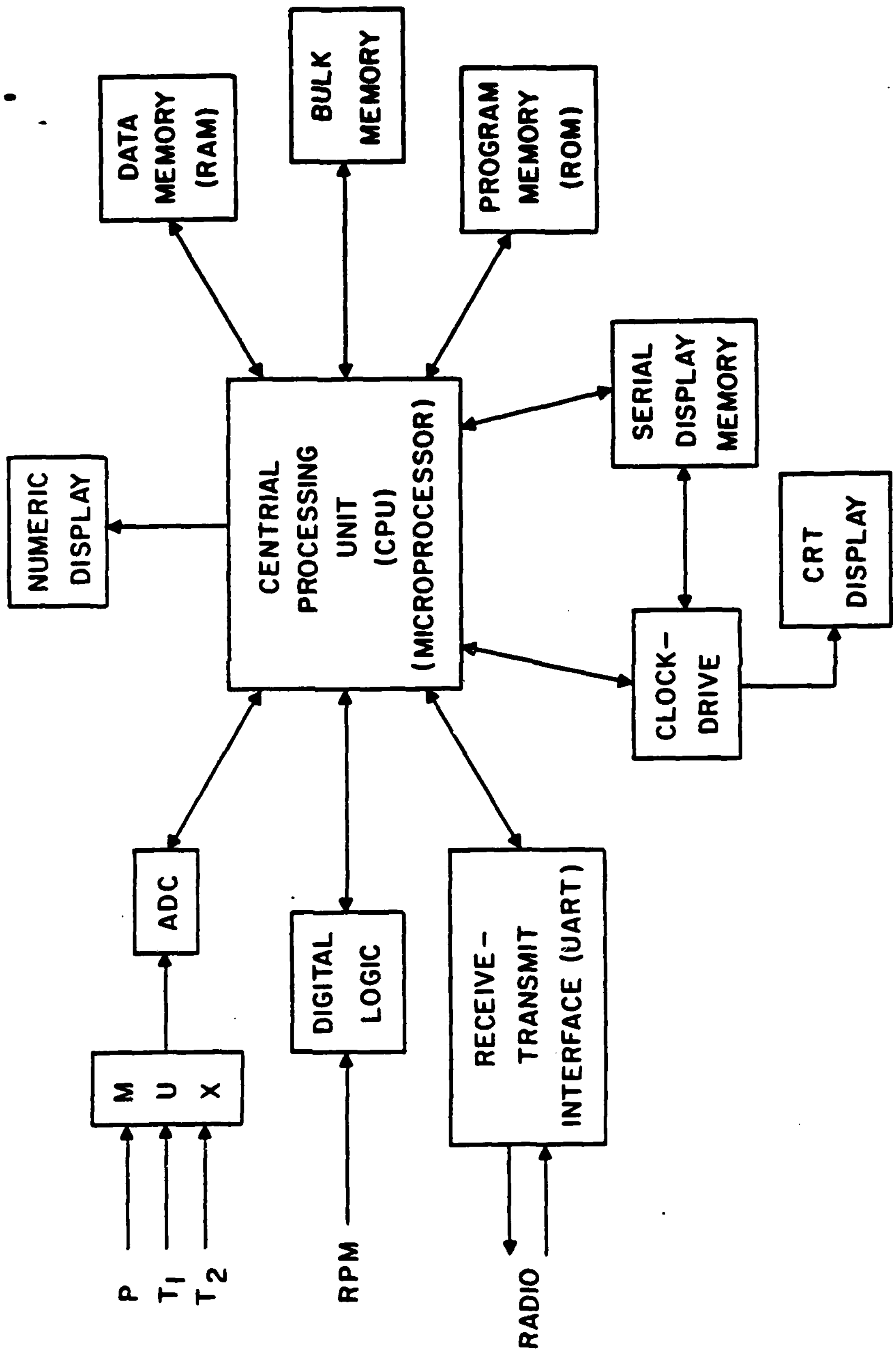


Figure 2.24

Response Times for the VLSI Instrumentation System

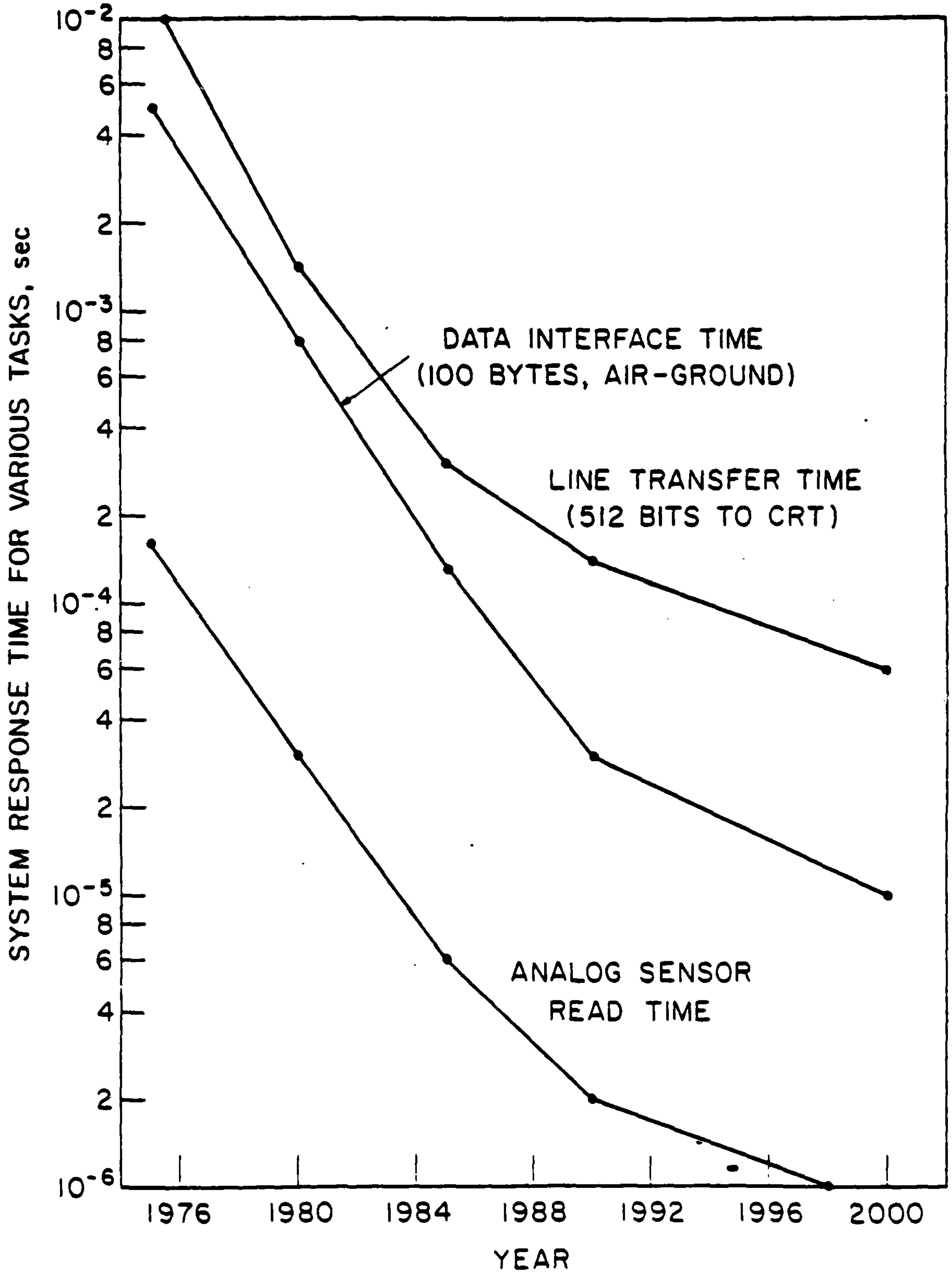


Figure 2.25

ACUMENICS

memory access, 1.5 msec of CPU time, 0.5 msec of data transfer, and 0.1 msec to access the CRT refresh memory. For a full screen change, 530 msec was required and was almost entirely attributed to data transfer, assuming this could occur at 1 M-bit/sec. It is noted that the bulk memory speeds here are optimistic. The 1979 Intel 7112 bubble memory has an average access time of 20 msec and a maximum data rate of 200 KHz. A faster data rate could probably be achieved using several smaller memories with outputs interleaved.

The data interface time is dominated by the CPU instruction cycle. Assuming a 1975 MOS instruction cycle of 2.5 μsec, 20 instructions per message byte, and 100 bytes per message, the data interface (formatting) time is estimated at 5 msec in 1975 and about 1.5 msec in 1980. Considering the expanded instruction sets and factor of three improvement in instruction cycles, progress in this area has closely followed the forecast curve.

The time to read and interpret an analog sensor is composed of an estimated 50 CPU instruction cycles and a 10-bit analog-to-digital conversion. In 1975, these were taken as 125 μsec and 40 μsec, respectively, for a total read time of 165 μsec. As with the other speed measures, this operation is expected to improve by a factor of more than 100 during this century.

The expected power of this instrumentation system is forecast in Figure 2.26. As shown in the figure, the microcomputer has a negligible contribution to system power, which is dominated by the bulk memory and the display. The 20 watts assumed for the 1980 bubble memory is in close agreement with the power requirements of the Intel IMB-100 board. It is noted that little progress is seen in decreasing system power requirements, largely because display power is not expected to change significantly.

System volume is shown in Figure 2.27, where again the contribution of the microcomputer is negligible. The most noticeable drop in size occurs with the adoption of a bubble memory and plasma display panel. Otherwise, the system size is nearly constant. It is noted that the system features could expand considerably in the same space during the forecast period since the microcomputer does not contribute significantly to the total system size. System weight closely parallels system volume (1).

System cost is shown in Figure 2.28 along with the relative size of its component parts. While the microcomputer is again a minor contributor, each of the other categories is significant. This study (1) assumed the estimated 1975 design costs of $200,000 would be recovered with the first 1000 systems. For telecommunications terminals, the design costs might be higher but so would the

Power Consumption Forecast for the VLSI Instrumentation System

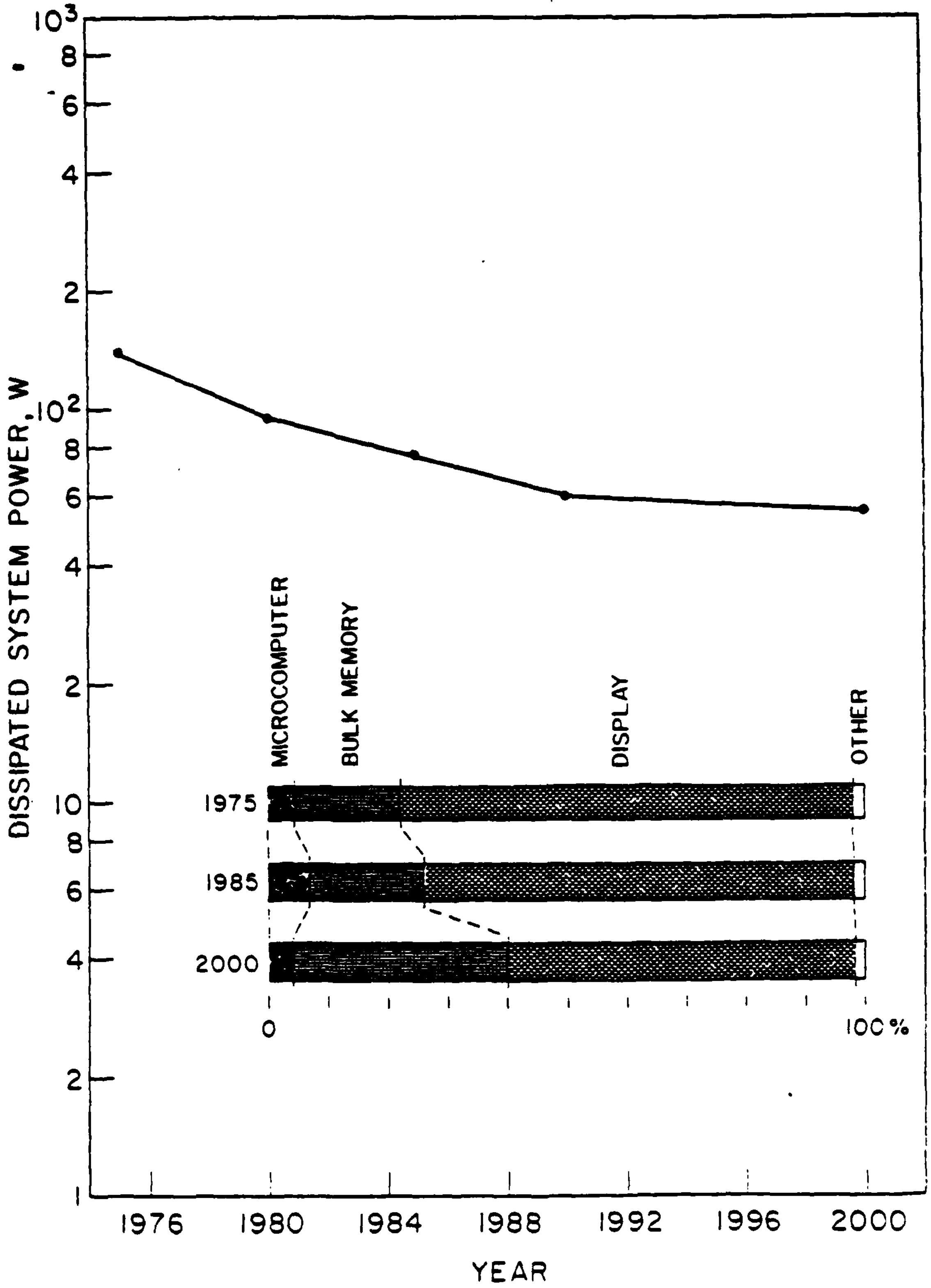


Figure 2.26

ACUMENICS

System Volume Forecast for the VLSI Instrumentation System

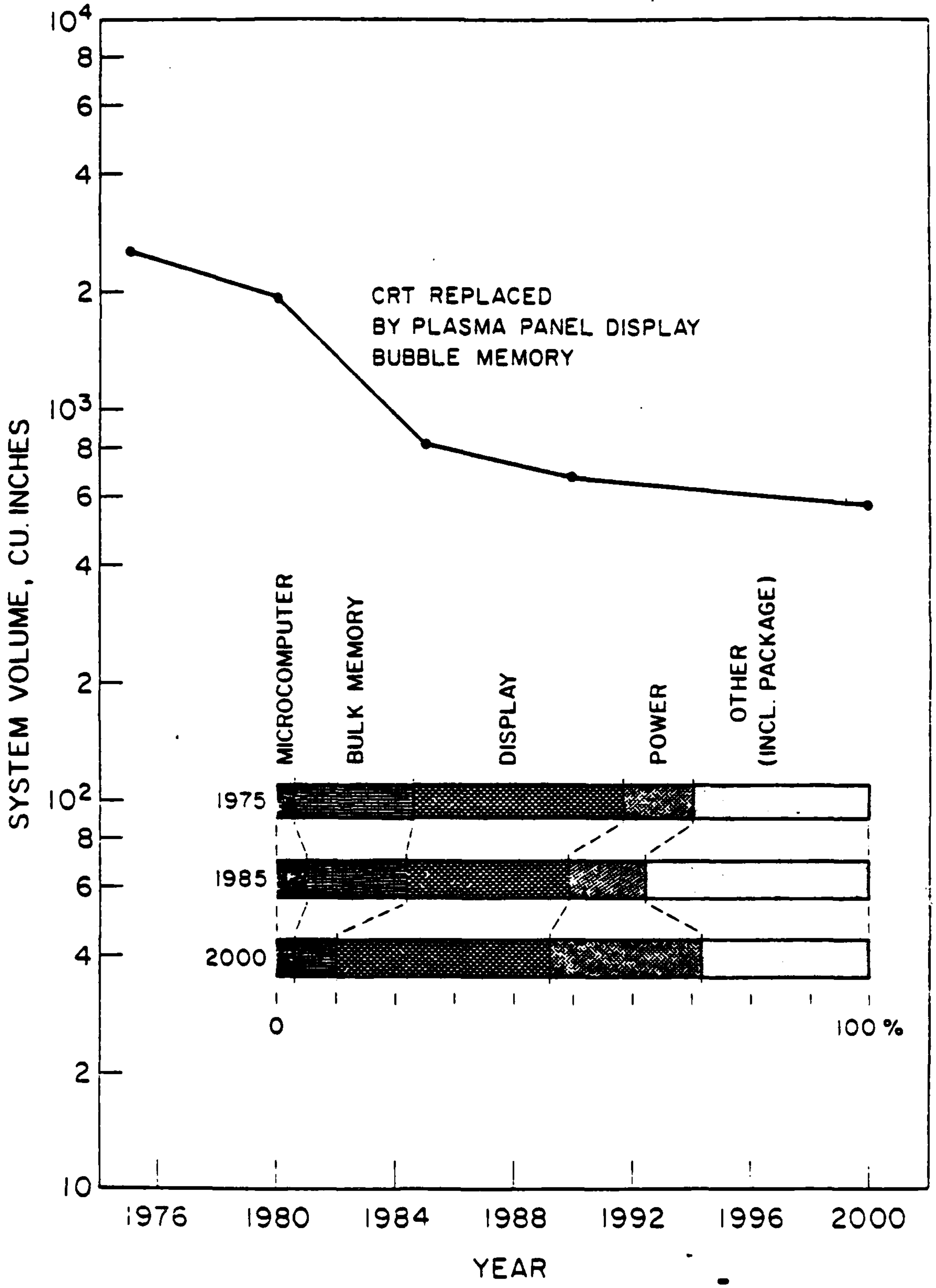


Figure 2.27

ACUMENICS

System Cost Forecast for the VLSI Instrumentation System

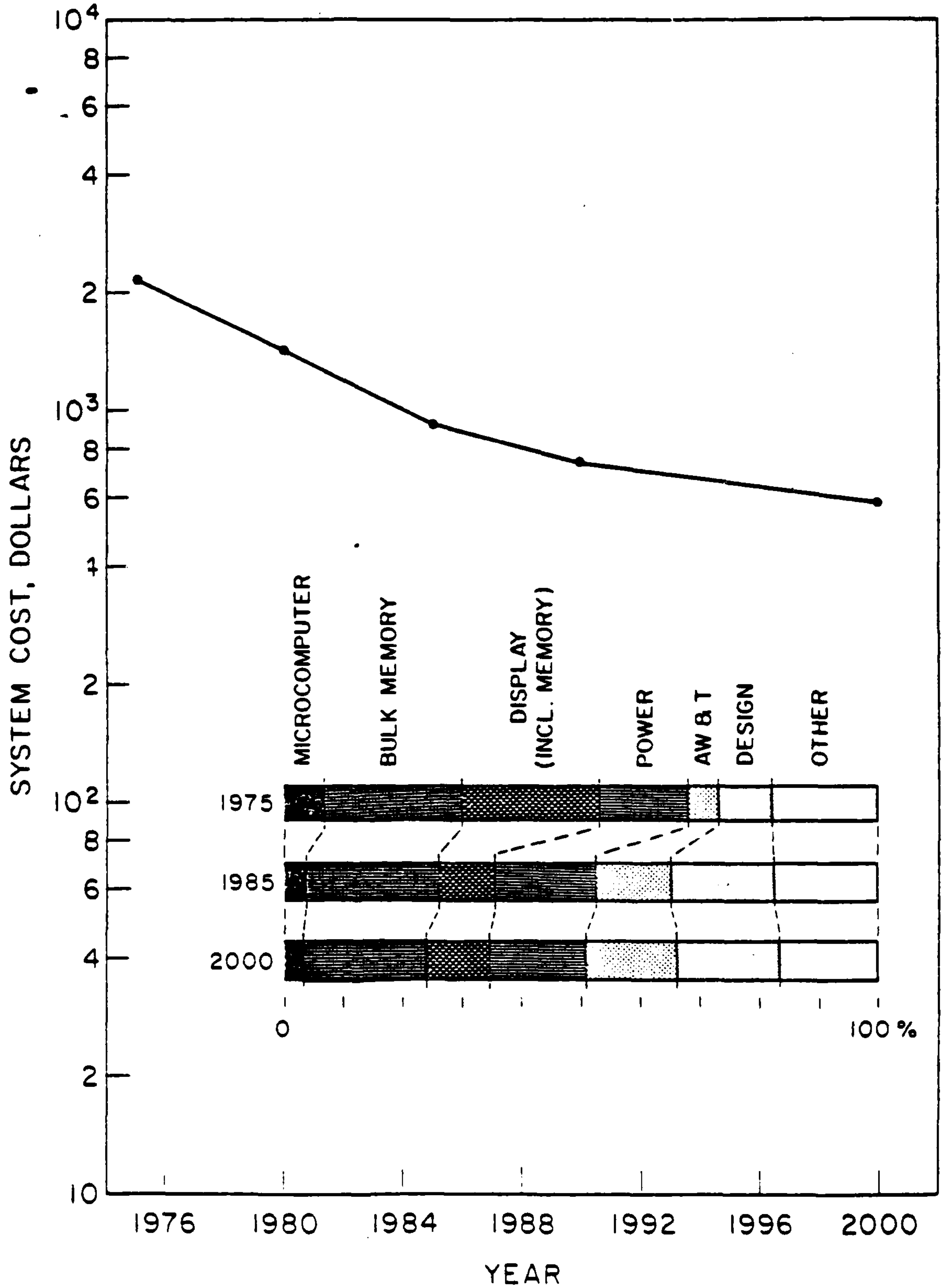


Figure 2.28

market, resulting in a possibly lower design component over the product life. While the cost of this system probably declines by no more than a factor of three or four over the next twenty years, the cost is dominated by peripherals (display, bulk memory, power supplies), and the system features could be expanded substantially with little change in cost.

As this study shows, the sophisticated capabilities of future instrumentation and input-output devices will be made possible by VLSI technology and microcomputers; however, the microcomputer will not contribute significantly to the speed, power, size, weight, or cost of the system, which will be much more dependent on peripheral devices, e.g., bulk memory and displays. While these devices are expected to improve substantially as the result of the exploding telecommunications market, they will nonetheless continue to pace the rate of progress in terminals and instrumentation systems.

2.7.2 The Impact of VLSI on Data Processing Systems

While instrumentation and input-output terminals are heavily dependent on many supporting technologies in addition to microcomputers, data processing systems lie at the opposite end of the scale, being heavily dependent on VLSI. Electronic switching systems are expected to follow the pattern of data processing systems closely, especially as time-division organizations expand. The only major exception will be in the line interface area, where some of the components will be slow to follow VLSI trends.

The data processing system designed in the previous work (1) was equivalent to an Amdahl 470 or IBM 370 machine, with approximately 200,000 gates in the CPU and 32 M-bits of read-write high-speed memory. Bulk memory and input-output devices were not considered in this study.

Speed in this system was measured in millions of instructions per second (MIPS), with approximately 160 gate delays per higher-level instruction. The speed forecast for the VLSI data processing system is shown in Figure 2.29. While in 1975, the VLSI system is slower by a factor of 20 when compared with large machines, by the late 1980s the VLSI system has surpassed the speed of 1975 machines. Were it not for power requirements, the processor could likely be realized on a single chip before 1990. System power is shown in Figure 2.30. CPU power is constant at 20 watts, while system speed increases. System power is determined by the high-speed read-write memory (RAM), which is assumed 25 percent active. Memory size is taken as less than the state-of-the-art by a factor of four and power is increased to reflect the emphasis on speed. For example, the 1985 system uses a 64 K-bit memory with a power per chip of 1.5 watts. The 64 K-bit chip should be widely-used industry standard by 1985, but power dissipation, even including the high speed requirement (effective CPU macroinstruction cycle is then 130 nsec), will likely be less than one watt per chip so that this power forecast is probably conservative by a factor of two to four. By the late 1980s, the VLSI power dissipation should be several hundred times less than the large machine of 1975.

In size and weight, the VLSI machine is significantly below the levels of present machines. Size decreases (1) from an estimated 16 cubic feet (1975) to 1 cubic foot by the mid-1980s, while system weight decreases in similar proportion, reaching 20 pounds by the mid-80s.

The cost of the VLSI data processing system is shown in Figure 2.31. Cost is dominated by the memory, which is expected to decline rapidly. The CPU cost, assuming high market volumes, could be less than $50 by 1990. Clearly, in the data processing area, large amounts of processing capability will be available in relatively small packages and at low cost. This will be of great significance to the design and organization of future electronic switching systems for use in telecommunications.

Speed in the VLSI Data Processing System

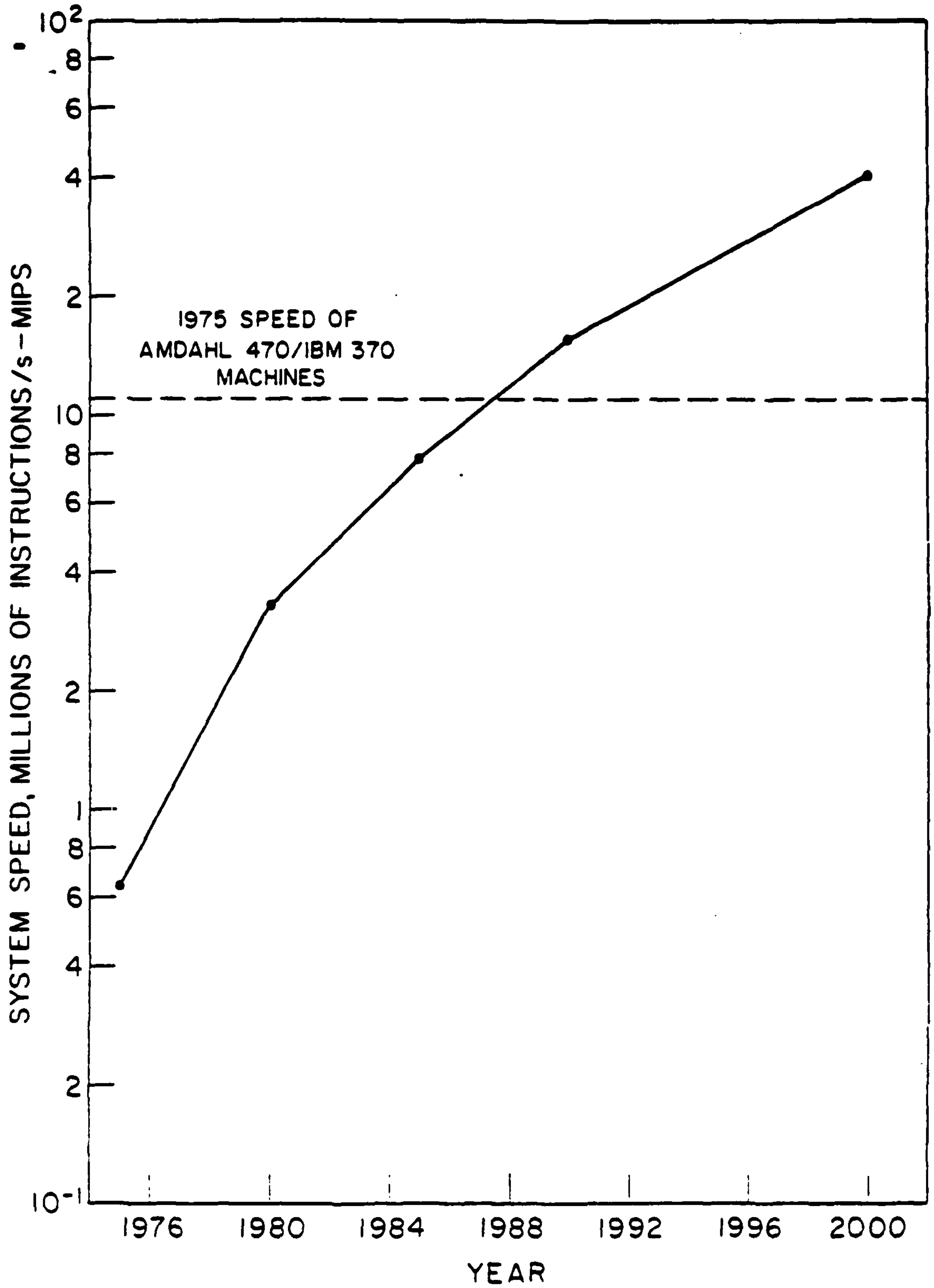


Figure 2.29

ACUMENICS

Power Requirements of the VLSI Data Processing System

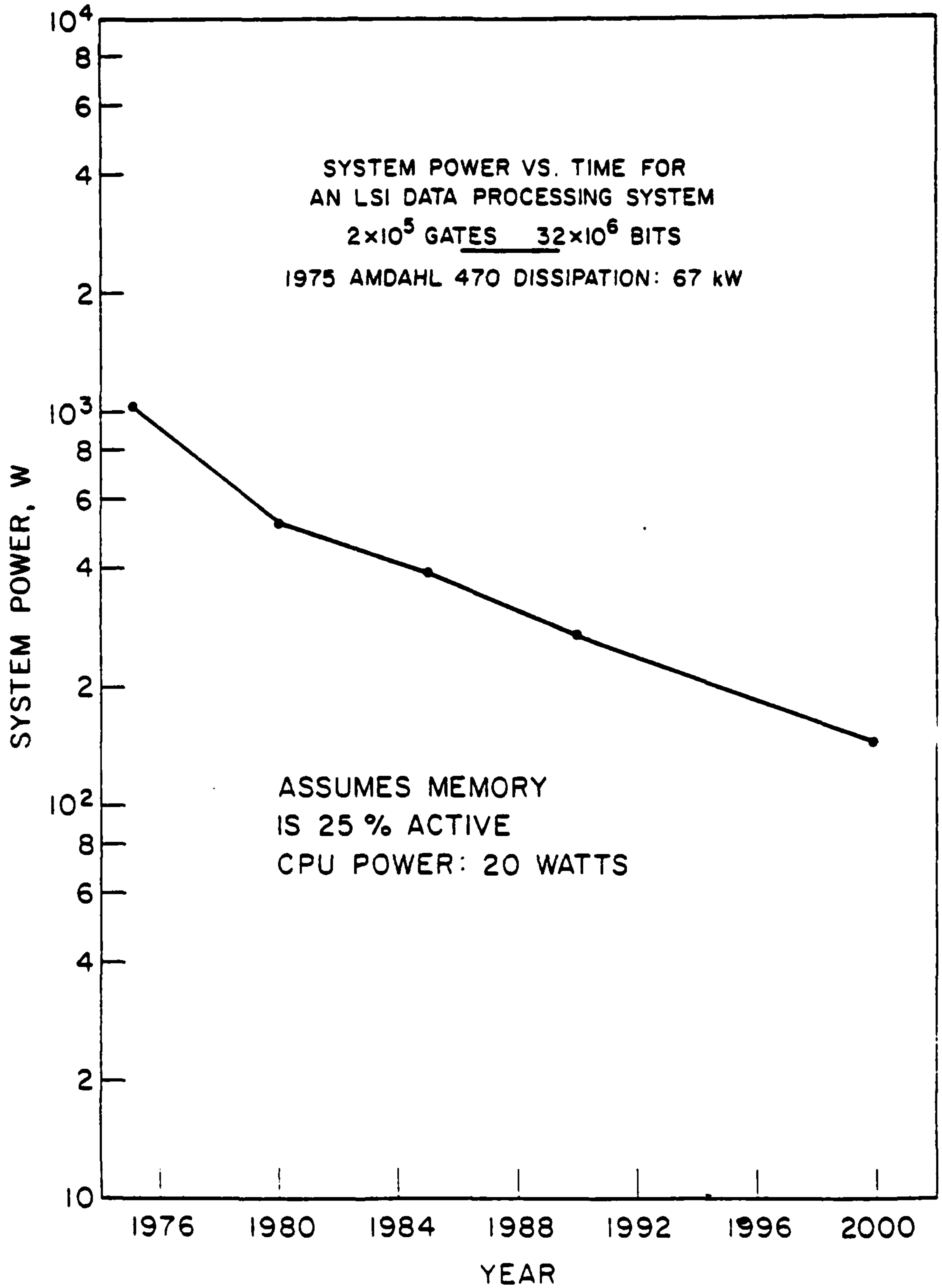
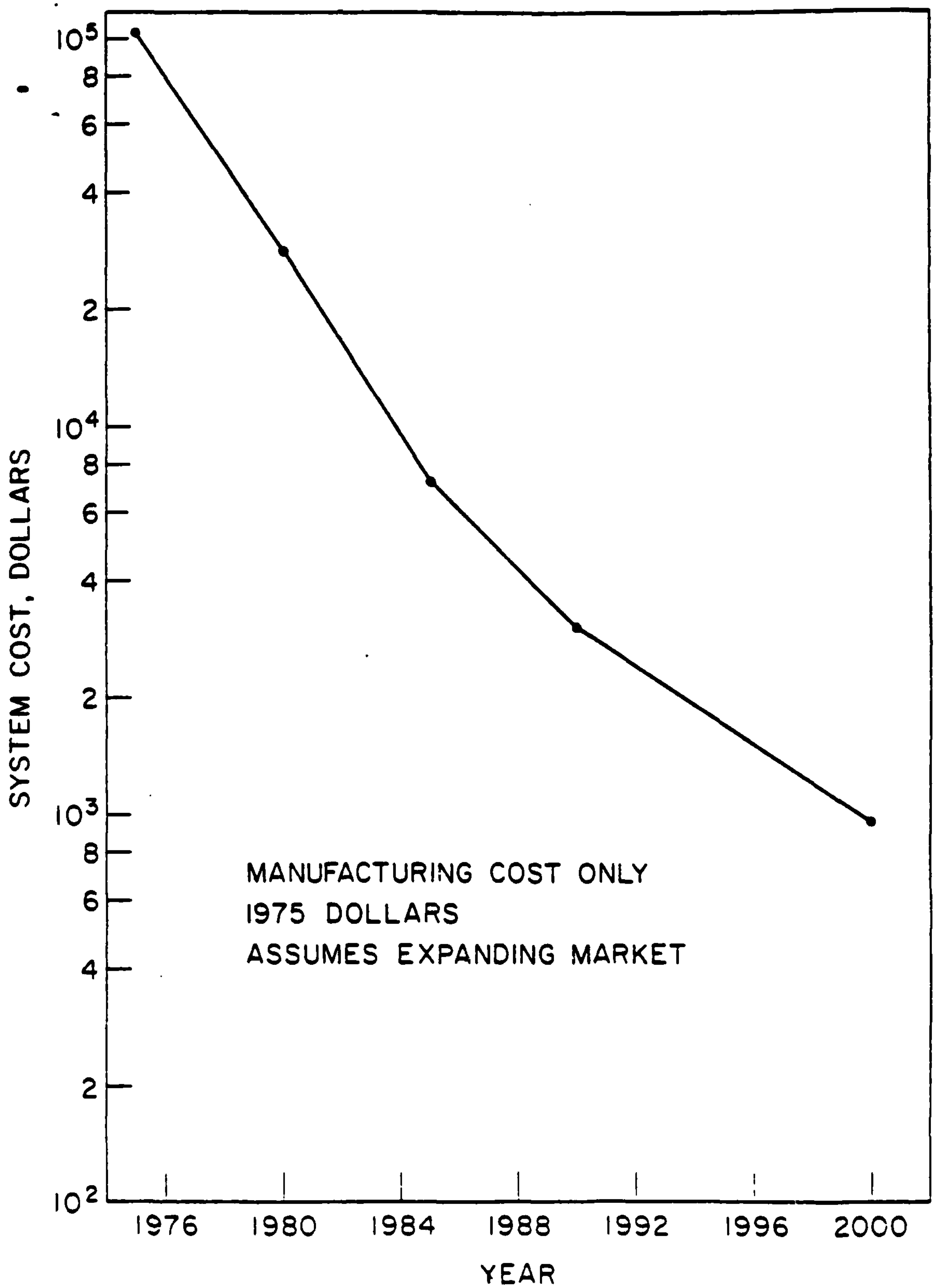


Figure 2.30

ACUMENICS



10^5
8
6
4
2
10^4
8
6
4
2
10^3
8
6
4
2
10^2
SYSTEM COST, DOLLARS
MANUFACTURING COST ONLY
1975 DOLLARS
ASSUMES EXPANDING MARKET
1976
1980
1984
1988
1992
1996
2000
YEAR

Figure 2.31

REFERENCES

1. K. D. Wise, K. Chen, and R. E. Yokely, Microcomputers: A Technology Forecast and Assessment to the Year 2000. New York: Wiley-Interscience, 1979.

2. A. B. Grebene, Analog Integrated Circuit Design. New York, Van Nostrand Reinhold, 1972.

3. A. S. Grove, Physics and Technology of Semiconductor Devices. John Wiley and Sons, Inc., New York, 1967.

4. D. J. Hamilton and W. G. Howard, Basic Integrated Engineering. McGraw-Hill, New York, 1975.

5. R. M. Warner and J. N. Fordemwalt, Integrated Circuits - Design Principles and Fabrication, New York, McGraw-Hill, 1965.

6. N. C. DeTroye, "Integrated Injection Logic - Present and Future," Iees Journal of Solid-State Circuits, 9, 206-211, 1974.

7. R. L. Horton, J. Englade, and G. McGee, "I^2L Takes Bipolar Integration Significant Step Forward," Electronics 48, 83-90, February 6, 1975.

8. K. D. Wise, "LSI, Microprocessors, and Electronic Automotive Control," Proc. Int. Conference on Automotive Electronics, IEEE No. 76CH1146-0 VT, pp 1.1-1.5, Sept. 1976.

9. P. W. J. Verhofstadt, "Evolution of Technology Options for LSI Processing Elements, " Proc. IEEE 64, 842-851, 1976.

10. J. Cunningham and J. Jaffe, "Insight Into RAM Costs Aids Memory-System Design," Electronics 48, 7-101-103, December 1975.

11. G. L. Schable and R. S. Keen, Jr., "Failure Mechanisms in Large-Scale Integrated Circuits," IEEE Trans. Electron Devices 16, 322-333, April 1969.

12. Intel Corporation Staff, Reliability Reports 7, 9, 10. Intel Corporation, Santa Clara, California. September 1975 - March 1976.

13. L. D. Anderson, ed., "Pattern Generation and Microlithography," IEEE Trans. Electron Devices (Special Issue) 22, July 1975.

14. D. E. Routh, C. E. Thompson, R. P. Mandel and J. W. Griffin, eds., "Semiconductor Microlithography," Proc. Soc. of Photo-Optical Instr. Eng. 80, June 1976.

15. E. Pease, ed., "Fine Line Device Fabrication," IEEE Trans. Electron Devices (Special Issue), 25, April 1978.

16. LB. Hoeneisen and C. A. Mead, "Fundamental Limitations in Microelectronics -- I. MOS Technology," Solid State Electronics 15, pp. 819-829, 1972.

17. B. Hoeneisen and C. A. Mead, "Limitations in Microelectronics -- II. Bipolar Technology," Solid State Electronics 15, pp. 891-897, 1972.

18. R. W. Keyes, "Physical Limits in Digital Electronics," Proc. IEEE 63, pp. 740-767, May 1975.

19. G. R. Madland, "The Future of Silicon Technology," Solid State Technology, pp. 91-95, August 1977.

20. R. M. Swanson and J. D. Meindl, "Fundamental Performance Limits of MOS Integrated Circuits," ISSCC Dig. Tech. Papers Feb. 1975, pp. 110-111.

21. K. Stein, "Noise-Induced Error Rate As A Limiting Factor For Energy Per Operation In Digital IC's," IEEE J. Solid State Circuits 12, pp. 527-530, October 1977.

22. I. E. Sutherland, C. A. Mead and T. E. Everhart, "Basic Limitations in Microcircuit Fabrication Technology," Report prepared for ARPA by the Rand Corporation, ARPA Order No. 189-1, November 1976.

23. R. Van Luyl, C. Liechti, R. Lee and E. Gowen, "4-GHZ Frequency Division with GaAs MESFET ICs," ISSCC Digest Tech. Papers, Feb. 1977, pp. 198-199.

24. R. C. Eden, B. M. Welch, R. Zucca, and S. I. Long, "The Prospects for Ultrahigh-Speed VLSI GaAs Digital Logic, IEEE Trans. Electron Devices 26, pp. 299-317, April 1979.

25. M. Klein, D. J. Herrell and A. Davidson, "Sub-100ps Experimental Josephson Interferometer Logic," ISSCC Tech. Papers, Feb. 1978, pp. 62-63.

26. Y. Akazawa, H. Kodama, T. Sudo, T. Takahashi, T. Nakamura, and K. Kimura, "A High-Speed 1600-Rate Bipolar LSI Processor," ISSCC Digest of Tech. Papers, pp. 208-209, February 1978.

27. H. Yoshimura, M. Hirai, T. Asaoka and H. Loyoda, "A 64-K-bit MOS RAM," ISSCC Dig. Tech. Papers, pp. 148-149, February 1978.

28. R. P. Capece, "Tackling the Very Large-Scale Problems of VLSI," Electronics 51, pp. 111-125, November 23, 1978.

29. J. M. Lee, J. R. Brewogel, R. Kunita, and C. Webb, "A 80 nsec 5V-only Dynamic RAM," ISSCC Dig. Tech. Papers, pp. 142-143, February 1978.

30. R. M. Jecmen, A. A. Ebel, R. J. Smith, V. Kynett, C. Hui, and R. D. Pashley, "A 25 nsec 4K Sttic RAM," ISSCC Dig. Tech. Papers, pp. 100-101, February 1979.

31. P. J. Salsbury, W. L. Morgan, G. Perlegos, and R. T. Simko, "High Performance MOS EPROMs Using a Stacked-Gate Cell," ISSCC Dig. of Tech. Papers, pp. 186-187, February 1977.

32. D. Toombs, "CCD and Bubble Memories," IEEE Spectrum, pp. 22-30, April 1978.

33. D. Toombs, "CCD and Bubble Memories," IEEE Spectrum, pp. 36-39, May 19178.

34. E. W. Pugh, "Magnetic Bubbles - Status and Future," ISSCC Dig. Tech. Papers, pp. 146-147, February 1978.

35. P. R. Gray and R G. Meyer, Analysis and Design of Analog Integrated Circuits, New York: John Wiley, 1977.

36. A. B. Grebene, Analog Integrated Circuit Design, New York: Van Nostrand Reinhold, 1972.

37. G. Kelson and H. H. Stellrecht, and D. S. Perloff, "A Monolithic 10-bit Digital-to-Analog Converter Using Ion Implantation," IEEE J. Solid-State Circuits 8, pp. 396-403, December 1973.

38. R. B. Craven, "An Integrated Circuit 12-bit D/A Converter," ISSCC Dig. Tech. Papers, pp. 41-41, February 1975.

39. J. J. Price, "A Passive Laser-Trimming Technique to Improve D/A Linearity," ISSCC Dig. Tech. Papers, pp. 104-105, February 1976.

40. D. T. Comer, "A Monolithic 12-Bit D/A Converter," ISSCC Dig. Tech. Papers, pp. 104-105, February 1977.

41. J. Schoeff, "An Inherently Monotonic 12 b DAC," ISSCC Dig. Tech. Papers, pp. 178-179, February 1979.

42. A. P. Brokaw, "A Simple Three-Terminal IC Bandgap Reference," ISSCC Dig. Tech. Papers, pp. 188-189, February 1974.

43. R. C. Dobkin, "Monolithic Temperature Stabilized Voltage Reference With 0.5 ppm/°C Drift," ISSCC Dig. Tech. Papers, pp. 108-108, February 1976.

44. R. A. Nordstrom, "High-Speed Integrated A/D Converter," ISSCC Dig. Tech. Papers, pp. 150-151, February 1976.

45. J. L. McCreary and P. R. Gray, "All-MOS Charge Redistribution Analog-To-Digital Conversion Techniques," IEEE J. Solid-State Circuits 10, pp. 371-379, December 1975.

46. R. H. McCharles, V. A. Saletore, W. C. Black, and D. A. Hodges, "An Algorithmic Analog-To-Digital Converter," ISSCC Dig. Tech. Papers, pp. 96-97, February

47. G. E. Moore, "Are We Really Ready for VLSI?" ISSCC Dig. Tech. Papers, pp. 54-55, February 1978.

48. G. E. Moore, "VLSI: Some Fundamental Challenges," IEEE Spectrum, pp. 30-37, April 1978.

Chapter Three

Input-Output Devices

3. Input-Output Devices in Future Telecommunications Systems

To the telecommunications user, the input-output device is the most visible portion of the overall communications system. It is his interface to the network and in a real sense defines the features he may expect from the system. To many users, it is the system, since improvements in switching and transmission are largely transparent to the casual user. Historically, input-output devices have evolved slowly. The most common of such devices is of course the telephone, which has changed very little since its birth in 1876. It was, and still is, a low-cost voice-only terminal of limited bandwidth. While its internal structure has evolved to become more compact and less expensive, its overall function remains unchanged. Improvements such as direct dialing and multifrequency signaling (e.g. Touch-Tone*) have evolved more as the result of central office (switching) needs than from pressure in the input-output area itself. The attempt to significantly expand the input-output (terminal) junction through the addition of two-way picture transmission (i.e., Picture-Phone*) was largely unsuccessful due to both cost and user acceptance. The addition of video to the audio communication link alone was simply not an attractive enough feature for most users to justify its cost, and for some it was an

*American Telephone and Telegragh Company

undesirable addition. Only in the business community was any market penetration achieved and even there, the service was perhaps premature during the 1960s.

During the 1970s, the situation in the terminal area began to change and change with increasing speed. The changes began in the business/professional sector as a direct result of the rapid advances in computer technology. With the proliferation of micro-computers, the number of intelligent machines to be communicated with exploded, and the number of terminal equipment manufacturers soared. These terminals were oriented toward data, not voice, and were principally for man-machine interaction. They combined a key-board, printer, input tape reader, and in many cases a cathode ray tube (CRT) display screen. Equipment designs have evolved rapidly and substantial reductions in size and cost have occurred. These terminals have not proliferated beyond the business community and into the home via the hobby market. As computer services have proliferated, the demand for data communications between distant machines or terminals has increased spectacularly. This communi-cations need has been met using the telephone network and an add-on modem (modulator/demodulator) to translate the data into a form suitable for transmission. In spite of relatively rapid developments in data terminals, the telephone has remained readily identifiable and largely unchanged as the principal interface with the telephone telecommunications network. However, this situation is now beginning to change and many are predicting an explosion of

new features in telecommunications terminals during the 1980s, which will merge the telephone and data set functions into super-terminals for both business and the home.

As the result of several regulatory decisions over the past 15 years, it has become relatively easy to interconnect an increasingly wide variety of devices to the telephone network. This network, while still referred to as a 'telephone' network, is already substantially involved in non-telephone traffic and hereafter will simply be referred to as the 'telecommunications network'. Some add-on devices, such as automatic answering systems, have already penetrated the home market substantially. In the future, however, major changes in the terminal area can be expected. First, it is unlikely that the telephone as we know it will disappear or change its basic form or function during the next forty years. It offers a basic service for which there will be a continuing need in both developed and developing nations. Extension phones, already a convenience in many homes, are also unlikely to disappear. Nevertheless, even the telephone is undergoing substantial internal changes. These redesigns are largely motivated by switching needs rather than anything in the input-output area itself. The trend toward digital switching and transmission has great advantages for both these areas and makes it desirable to translate voice information into a serial bit stream at the telephone so that (class 5) end offices can go digital along with the toll network. Also, the power and ringing requirements of the familiar '500-set' telephone

have been a major factor preventing the adoption of low-cost solid state technology in local switching offices. The development of a solid-state tone-ringing telephone is underway and its deployment in new systems is likely during the 1980s. It will have more impact in the switching area than any other single development. In the terminal area, however, the telephone will not appear radically different in size, performance, or cost.

In parallel with the development of the solid-state digital telephone, revolutionary developments in much more diverse communication centers for business and the home can be expected. These terminal centers will be modular and will represent a merge of communications, control and entertainment functions, only a few of which are now available individually (1-3). Such centers will combine television, home and commercial movies, games, newspaper and library access, mail, radio and recorded music, environmental control, security, financial and other record keeping, funds transfer, and general computing functions. How many of these features will find widespread application in the home during the next forty years will depend on developments in a wide variety of areas, but certainly many of them will. Television, radio, and stereophonic sound are already widespread on an individual basis. Video games and cassette-based movies are available and starting to effect major market penetration. Home hobby computers are powerful and efficient for many applications and home calculators, now common, are growing in sophistication and features. So far,

most of these features are independent of each other. As more and more features become desired by an increasing market segment, however, some will naturally combine (e.g., television, video cassettes, video games). Others will become modular add-ons to a core facility so that some portions of the facility can be shared. Probably, the market will support both modules and stand-alone systems as it has in the stereo equipment industry. With the growing sophistication of features and equipment, the development of standardized interface protocols will be of major importance.

Control functions such as environmental control, power management, and home security are dependent on the home computer, a variety of peripheral devices, and probably on an interface to the broader telecommunications network for billing, remote control, and other functions. In an increasingly energy-conscious society, there will be substantial motivation to develop such systems on the part of both the utilities and individual users. Such procedures have already been designed into some new office buildings, and over the next forty years it is likely they will permeate most, if not all, homes. These systems will conserve energy and make energy distribution more predictable and manageable. substantial policy questions over the degree to which control is taken out of the hands of the user will have to be settled.

Many of the more significant communications developments for the home will require substantial amounts of information flow over

the telecommunications network, either broadcast or wired. Newspaper and library access as well as electronic funds transfer, electronic mail, teleconferencing, and remote shopping are in these categories. These features will likely exert a significant impact on travel and energy usage, and are of direct concern to organizations such as the FAA. Most of these functions are being seriously considered or developed at this time, and a few are available on a very limited scale. Their widespread deployment before 2020 is likely, although substantial problems remain to be overcome. The codification of information in an electronically retrievable form, even if only on retrievable microfilm, is a job which could extend beyond the 21st Century.

Several information retrieval systems are in limited service today, the most advanced being teletext and viewdata systems. Teletext is a broadcast system in which the digitized information is transmitted on two normally unused lines of the vertical TV blanking signal. Information is cyclically repeated. The television receiver grabs the page of interest, stores it locally, and subsequently displays it on command. The information is grouped into a limited number of specific categories from which users select by entering a numerical code on a keypad. The system is basically non-interactive, and since the number of unused transmission lines in the television transmission format is limited, so is the total data content. Nevertheless, a considerable amount of information could be made available using this approach. Twenty-five

pages of a standard newspaper require less than 0.5 seconds to transmit at TV broadcast rates so that even using only two lines per frame, the cycle time (user access time) for a block of data this size is on the order of a minute or less. The system is relatively simple, using already-available transmission facilities for television, the available television display, and allowing the user considerable control over the capabilities of his system via the display size and memory size for page of storage purchased. A variety of user-selected options (e.g., hard copy) are feasible for such systems, which are limited in scope but a likely replacement for newspapers before the 21st Century.

View-data systems rely on digitally-stored information, organized as pages, which are modulated onto an audio-frequency carrier and transmitted over a telephone channel. The user indicates the information desired via a numerical code. The view-data system is basically interactive, with the range of information available limited only by the size of the database. The system is wired and subscriber requests are transmitted at 75 bits per second (bps) in most present systems (2). Data is sent to the subscriber at 1200 bps so that a page of information can be accessed in a few seconds. The view-data approach is more general than teletext, but its broad usage is clearly dependent on the capabilities of the broader telecommunications network. The eventual system used for telecommunication, electronic mail, and electronic funds transfer in the 21st Century will likely be a merge of the teletext and

view-data approaches with general information of interest to many subscribers broadcast and personal information wired. The realization of local fiberoptic transmission facilities will play an important role in the extension of viewdata from text-only into personalized video displays for two-way communication, shopping, and other applications.

The general communications/entertainment terminal center of the future can be divided into a number of specific sections as shown in Figure 3.1. In the control complex, a sophisticated microcomputer, equivalent to a large minicomputer of today, will direct the system. This computer will be realized using VLSI technology and will utilize a non-volatile program memory (ROM) containing from 100,000 to a million or more bits of dedicated code, a high speed random access memory (RAM), and a bulk storage memory of slower speed but having many megabits of storage capacity. One or more codecs (coder-decoders) and other information-formatting chips will accomplish the analog-digital conversion functions and satisfy the network interface requirements. A variety of intelligent sensors and actuators will also interface with the control complex for purposes of environmental control, power management, and security.

The input and output sections employ many similar media. Both interact with the network over a two-way data link, which should probably be more properly termed an information channel since it

Elements of a Communication/Entertainment Center of the Future

INPUT FUNCTIONS	CONTROL FUNCTIONS	OUTPUT FUNCTIONS
Remote data link (network interface)	Microcomputer	Remote data link (network interface)
Local Voice	Program memory	Voice and Music
Local Magnetic Store	Bulk Storage	Local Magnetic Storage
Scene and Graphics (Video)	Codec	Video Display Screen
Keyboard entry	Specialized interface devices	Printer (hard copy)

Figure 3.1

will handle voice as well as data. Additional inputs will be derived from magnetic tape (data, voice, or video), a keyboard, a live video camera and local voice. In the long term, voice will not only be transmitted directly, but will also be recognized and interpreted in the terminal itself to allow a variety of voice commands. The output modes will involve sound and visual (CRT) display as well as more permanent media such as tape and printed copy. While most of these media are available today, the cost of such facilities is prohibitive except in a relatively small number of installations where it is possible to share the equipment over many users. The major change over the next few decades in terminal equipment will be that as network-derived services expand, the cost of such equipment will decline to the point where it will be affordable in the home.

In the remainder of this chapter, several key areas of significance in terminal equipment are examined in detail. Expected progress in this area is then expressed in quantitative terms.

3.1 Input-Output Terminal Control

Most of the components needed to control the communications center shown in Figure 3.1 were discussed in detail in Section 2. Microprocessors costing less than fifty dollars and capable of controlling such a communications/entertainment complex are probably available today and will certainly be available and less expensive by 1985. Semiconductor memory (ROM and RAM) for

program and high-speed data storage is rapidly increasing in density and should be available at 5.2 K-bits/chip (ROM) and 256K-bits/chip (RAM) in 1985, with a cost of 10 cents/K-bit or less. The most promising candidate for non-volatile mass storage of information requiring relatively fast access is the magnetic bubble memory, and megabit chips are available today. By 1985, higher bit densities should be available and memory cost, which is still relatively high (more than 100 cents/K-bit) should have fallen between one and two orders of magnitude. It is noted that actual memory prices reflect the balance between supply and demand and could fall more slowly if the present supply shortages relative to demand persist. For the general control area, however, it is virtually certain that the VLSI chips needed for future communications terminals will be available before 1990.

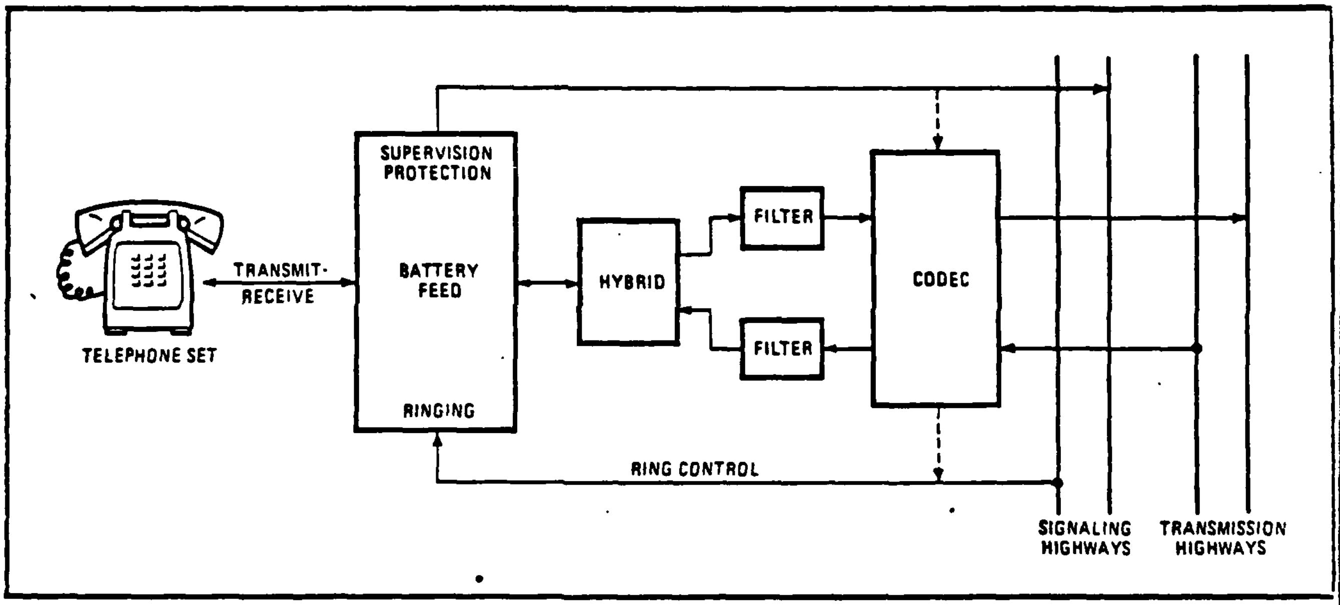
In addition to the microcomputer itself, a variety of special purpose chips implementing functions unique to telecommunications will be required as part of the control complex. These chips will handle information formatting, interface protocols, some filtering, and other functions. They will require a mix between analog and digital circuitry, high integration levels, high precision, and probably high speed. Many of the needed functions are now being integrated, and the first round of resulting chips has been described (1,4-6). Of the many needed functions to be realized in monolithic form, probably the most important to the development of digital input-output devices is the codec. These devices have

recently been the focus of widespread development efforts (6,7) and since they represent a blend of the high precision analog and digital requirements characteristic of future telecommunications systems, they will be used as an example and discussed in greater detail.

Codecs are basically two-way analog-to-digital/digital-to-analog converters. They are needed to convert the basic analog speech format into a digitally encoded series of bits. Since toll transmission over trunks was the first portion of the network to go digital, codecs first found application in channel banks such as the Bell System D3, which interfaces with the network. These codecs are relatively expensive and are therefore shared over several lines. With the realization of monolithic codecs which meet the high performance specifications of the D3 channel bank, costs have come down to the point where codecs are being applied in local switches and PABXs (private automatic branch exchanges) and are seriously being considered for inclusion in the telephone itself.

Figure 3.2 shows the codec as it might be applied in a PABX or central office (CO). The telephone interfaces with the network through a line interface circuit which provides supervision, battery feed, and ringing access to the telephone. The analog voice signal is passed by the bidirectional line interface to a hybrid which splits the transmission path into two unidirectional

Use of a Monolithic Single-Chip



Codec in a Digital PABX or Central Office (7). The codec and other components shown may evolve into the telephone itself, permitting digital local subscriber loops.

Figure 3.2

ACUMENICS

paths. In the transmit direction, the voice signal is first band limited by a low pass filter with a sharp cutoff at 4 kHz, which is known to permit sufficient voice fidelity. As shown by the Nyquist sampling theorum, any bandlimited analog signal can be reconstructed in its entirety if sampled at a rate at least twice its highest frequency. Accordingly, the codec samples the filtered voice signal at an 8kHz rate (one sample every 125 μsec). The codec performs an 8-bit analog-to-digital conversion on each sample and inserts the encoded sample into a time slot in the digital transmission bus. In the receive direction, the process is reversed. A digital sample is extracted from the transmission bus and an 8-bit digital-to-analog conversion is performed. The quasi-analog sample is smoothed by a second filter and fed to the user.

Figure 3.3 shows one design for a single-chip in greater detail. A single DAC is time shared between the encode and decode functions, operating as part of a succcessive approximation ADC in the encode mode. With an allowed sample interval per channel of ony 125 usec, ADC conversion time is no problem if the successive approximation technique is used and the codec is used on a dedicated per line basis. Of greater concern, however, have been the cost and power dissipation of the per line codec. With the continued development of CMOS designs, per line codecs should be very low in power and economically practical for use in the input-output device itself. Shared codecs allow cost and power to be

Block Diagram of a Single-Chip Codec (1)

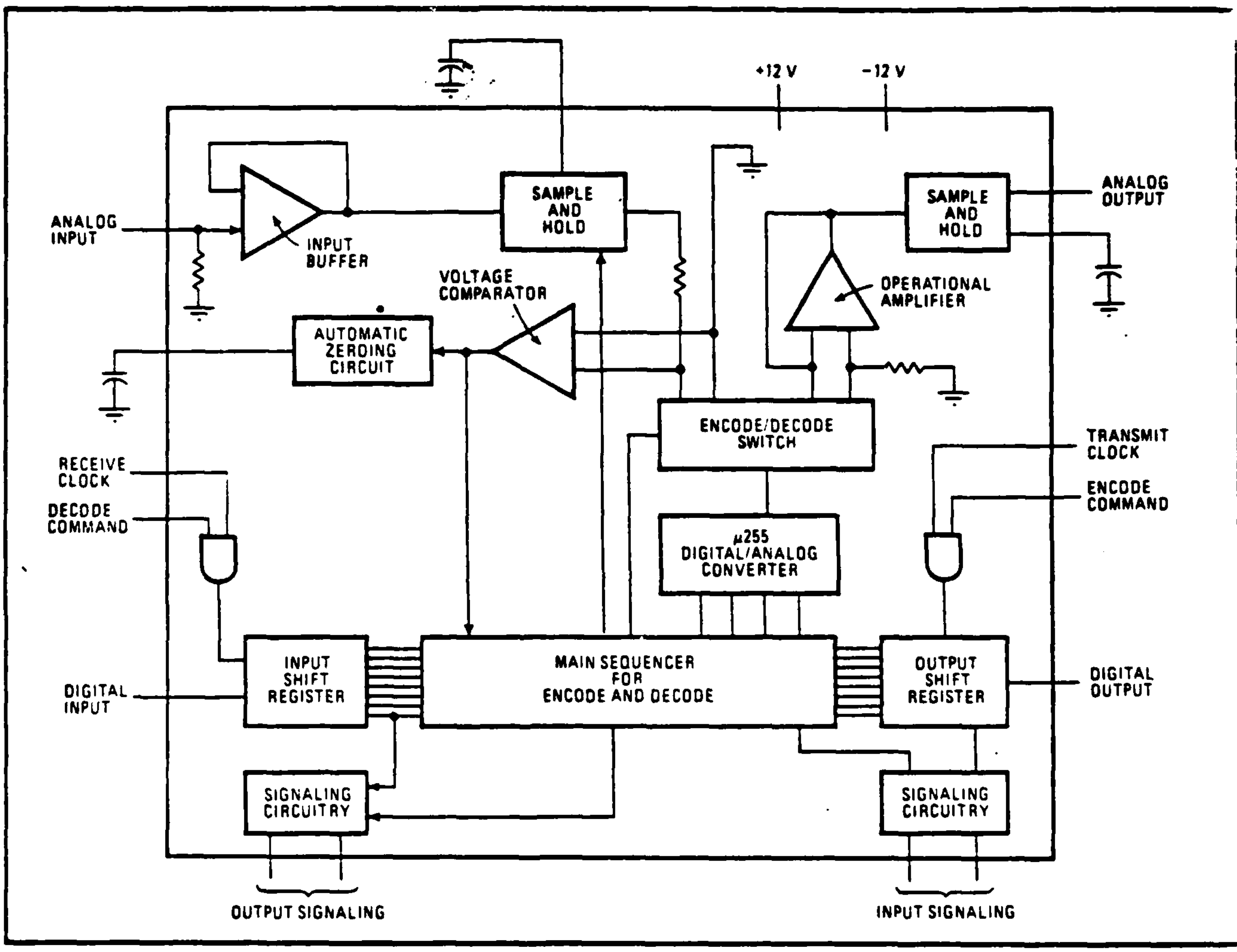


Figure 3.3

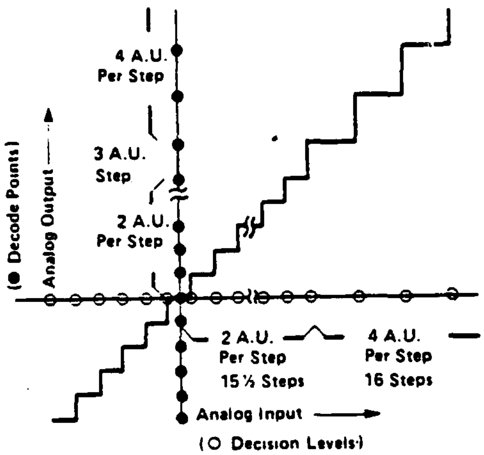
divided over the number of lines served but also require that the effective sample interval be reduced. For a codec shared over 24 voice channels, the ADC conversion time would have to be less than 5 μsec, which is too fast for present technology but should be available by 1985. If codec economics alone were the deciding factor, then the shared approach would be preferred; however, the advantages in extending digital transmission to the local subscriber loops may justify the per channel approach, particularly in conjunction with fiberoptics.

The sampled voice signals are coded in the pulse-code-modulated (PCM) format used by the transmission network. The encoding is not linear but corresponds to a piecewise-linear (segmented) approximation to the analytic μ-255 logarithmic companding law (8). Two approximations are recognized -- the segmented μ-255 law and the segmented A-law. The μ-law is used in North America, while the A-law is used in Europe. Differences are relatively minor and most codec manufacturers provide both types. Both laws approximate the logarithmic curve by a series of eight segments on either side of the origin, with each segment having sixteen equal steps. For μ-law devices, the step size doubles in size from one segment to the next going away from the origin, while for A-law devices this is modified slightly in that the first two segments on either side of the origin have the same step size (7,9). Thus, the resolution of the conversion is high for low-amplitude signals but is reduced at high amplitudes. The signal amplitude

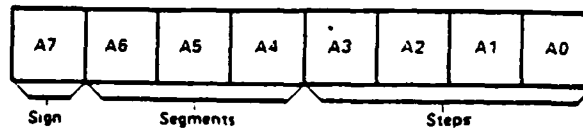
is encoded in an 8-bit PCM format consisting of a sign bit, three bits representing the segment, and four bits representing the number of steps within that segment. Figure 3.4 summarizes the commercial PCM formats used.

The codec is the key element needed for the realization of an all-digital network. As Figure 3.3 shows, it can be realized by combining digital logic with one or more DACs to allow the encode/decode functions to be performed. In terms of speed, the ADC process is limiting, and the forecasts provided in Section 2 relative to ADC speed are applicable to codecs. While the accuracy of the ADC is relatively low for high signal levels, for low signals the required resolution in the first segment is comparable to 12 or 13 bits for a linear converter. This is achievable in present monolithic designs. Codec cost will be influenced by its complexity and its volume. Most present die sizes are in the 20-35,000 square mil range (7), and future designs should reduce this substantially. For the basic codec function, where volumes should be very high, chip cost can be expected to follow the curve shown in Figure 2.23, possibly dropping below $5 by 1985. It is likely that the codec chip will expand to include the input and output filters, and designs using switched capacitor, CCD, and other filtering approaches are being explored. These filters require sharp cutoffs and very small passband ripple and are a challenge in themselves; however, it is likely that monolithic designs can be realized which are less expensive then present active filters

PCM Transmission Formats (2,9)



An illustration of the segmented μ-255 coding law near the origin. The smallest segment on each side of the origin contains only 15.5 intervals, while all other segments contain 16 equal intervals (steps). One analog unit (A.U.) is the magnitude of one half-step in the smallest segment.



Binary format of an 8-bit companded PCM word.

Commercial PCM Digital Formats

	North American	European
Sampling	8 kHz	8 kHz
Bits/channel	8	8
Channel Capacity	64 kb	64 kb
Channels/frame	24 (24 voice + 0 signaling)	32 (30 voice + 2 signaling)
Bits/frame	193 (192 + 1 for framing)	256
Encoding	μ-law	A-law
Signaling	Borrow LSB every sixth frame	Separate chann (0 and 16)
Transmission mode/rate	bipolar/1.544Mb/sec	bipolar/2.048 Mb

Figure 3.4

ACUMENICS

(about $7 each), and by combining the filters with the codec, the resulting chip will be attractive for per-line use. As a point of comparison, the common 500-set telephone has a cost in the range of fifteen dollars, so the cost sensitivity in a simple terminal such as this is extreme, and the cost of new designs must be viewed considering the overall system. For a more extensive communication/entertainment center, the cost of the codec will be relatively small. By 1990, all of the components for the control section of such a center will be available and consistent in price with applications in the home. Cost will likely be dominated by peripherals such as displays and by the hardware associated with other features such as speech recognition. These aspects of input-output devices are discussed in the following sections.

3.2 Displays for Communications Terminals

One of the most important output devices for use in telecommunications equipment is the visual display. Such displays will remain the dominant mode of presenting complex information to the user. While the lowest-cost devices may remain limited to alphanumeric formats, it is likely that fully graphic displays of at least television quality will be the goal for most communications systems. In this section, the competing display technologies at the present time are first described individually, followed by specific comments relative to character and imaging displays.

3.2.1 Display Technologies

The display area has been extremely active during the past decade with a number of competing technologies in exploratory development. The diversity and number of these approaches suggests some fragmentation of effort and the fact that no single approach has been very effective in meeting the full range of display needs. For aircraft and air traffic control (ATC) applications, there is a need for both character (alphanumeric) and imaging displays. Important characteristics include high reliability, low cost, small size, light weight, low input power, easy interfacing with digital drive circuitry, low luminance noise, and suitable brightness under varying conditions of ambient lighting. Some of the technologies under present development are primarily for character displays while some are for panel imaging. Since the requirements in these areas differ, they are discussed separately following a description of the present competing technologies. For more comprehensive, in-depth discussions of displays, the reader is referred to several recent reviews (10-14). It should be emphasized that the aircraft market for displays will not have a significant impact on the development of display technologies, which will be oriented toward entertainment (including video games) and the broader communications market.

The Cathode Ray Tube (CRT)

Although the CRT is among the oldest of the display devices, it is unmatched in cost and quality for many applications and

remains the standard to which all other displays are compared. The CRT is a vacuum device in which an electron beam is deflected over a screen coated with a luminescent phosphor. The beam deflection, which is controlled by the input video signal, can be operated in either a raster-scan or random-access mode. In raster scan, the beam sweeps the entire display area line by line. This mode is preferred for complex displays and TV-type imaging. The random-access mode directs the beam only to those parts of the screen where information is to be written, and is preferred for limited graphics and for character displays. In either case, the high peak brightness and relatively long persistence of the phosphor permits a high average brightness at scan rates of 1/30 second or faster. This is important for cockpit displays, where ambient light can approach 10^5 lm/m^2. Tube phosphors have uniformity and good dynamic range. Both black-and-white and color tubes are widely available, and a great deal of work on each tube type continues. Efforts in black-and-white CRTs have recently concentrated on improving deflection sensitivity and on improving the writing speed. For high-resolution applications, tubes having diameters of 11 inches and resolution exceeding 40 lines/mm are available (15).

For airborne applications, the advantages of the CRT include widespread availability, high brightness, relatively simple drive requirements, and low cost. Its disadvantages include its relatively large size and weight, and the high voltage required for

its operation. The CRT has a strong place in display technology and handles most jobs very well. In spite of developments in alternative technologies, the CRT continues to supply tough competition and is expected to be widely used for another decade or more.

The Light Emitting Diode (LED)

Next to the CRT and perhaps the incandescent light, the LED is probably the most familiar and widely used display device. The diode here is formed in gallium phosphide (GaP), gallium arsenide (GaAs), or gallium arsenide phosphide (GaAsP) crystals formed epitaxially. Red LEDs were the first to be widely available, but have now been joined by yellow and green devices. Recent work has centered on techniques for achieving improved efficiency and alternative colors. In GaP, a major focus remains on the development of nitrogen-doped diodes for green-yellow emission, resulting in the maximum contrast as perceived by the human eye. Efficiencies, reproducibility, and uniformity are improving, and some progress using GaN diodes emitting in the blue has been reported.

A major new research area involving LEDs with direct relevance to telecommunications is that of devices emitting in the infrared with wavelengths near the 1.3 μm low-loss transmission window for optical fibers. Most recent activity has centered on the use of quaternary III-V alloys to achieve independent control of the energy bandgap and lattice constant in the emitting structure. This

control is essential for high efficiency, long-lived, double-heterojunction LEDs, which are typically grown by liquid phase epitaxy and consist of stacked layers of GaAsSb/GaAlAsSb (16) or GaInAsP/InP (17). Such devices have recently achieved output powers of 54 mw and an external quantum efficiency of 27 percent at 0.94 um (16). Lensed 18 um-diameter devices have achieved launchpowers of over 0.2 mw into optical fibers at 1.3 um and have exhibited optical bandwidths beyond 300 MHz (17). Work in these non-display aspects of LEDs is discussed further in Section 5.

LEDs continue to be wide ; used in the calculator market in spite of the relatively high current levels needed to produce acceptable intensity. Several millamperes per segment are typically required. The diode voltages required are in the range of a few volts or less so that LEDs interface well with many integrated circuit technologies, especially bipolar. Material quality is increasing, and costs are decreasing so that this technology promises to continue to enjoy wide use in character displays for some time.

Gas-Plasma Devices (GPD)

These devices have been undergoing intense development and are probably the most serious competition for the CRT in large panel displays. The device consists of parallel front and back plates, generally less than 10mm apart, containing orthogonal

conductors. The space between the plates is typically neon mixed with an inert gas such as nitrogen. Both ac and dc operating GPD modes are possible.

Gas plasma discharge displays employ different techniques than the CRT. In a CRT, an electon gun sweeps back and forth, energizing phosphor dots on the screen. In the GPD, the "dots" are the intersections of a grid of conductors etched into two glass plates. When two crosswise conductors are energized, a high voltage of 150-200 volts is applied between the electrodes and the voltage at their intersection creates a bright dot of gas plasma. These dots are used, through matrix addressing techniques, to form alphanumeric displays.

The GPD screen is also substantially more compact than a comparably sized CRT. The commonly used 12 inch, 2,000 character CRT screen requires a terminal that measures 14 to 16 inches deep, and the depth must be increased to increase the screen size. The same sized GPD, with capabilities of up to 4,000 characters, requires a depth of only 1 1/2 to 3 inches. In addition, the depth required to store the GPD electronics is not linked to screen size. Therefore, the GPD has significant advantages when space available is constrained.

Another advantage of the GPD to the CRT is the lack of "flicker" that can cause eye fatigue after 1 1/2 hours of continous use. Because the image on a CRT is made by an electron beam scanning the screen, the image begins to fade as the beam passes. The flicker occurs when the image is replaced on the next scan. This flicker does not occur with GPD's because of the matrix addressing techniques employed.

Current ac GPD technology employs up to 512 x 512 matrix elements with resoultion of 60 lines per inch and a range of 2-4 thousand characters. Although most displays are red or orange in color, blue and green have also been reported. The largest ac panel developed on an experimental basis for military use has been a 24-inch-diagonal unit that can display over 21,000 characters within its 1,024 x 1,024 element addressable matrix, at a resolution of 60 elements per inch. These panels are expensive, however, because of the extensive driver mechanics required.

The major cost item in the ac GPD has been the need to address each grid intersection individually. This situation can be changed through the use of a new integrated circuit designed by Texas Instruments. Use of this integrated circuit can potentially reduce costs because each circuit can address 32 conductors instead of only one. This would reduce the number of printed-circuit boards required to drive the display panel as well as the associated

electronics. This 32-electrode controller has already been shipped to some customers in prototype quantities, and the addressing of larger numbers of conductors with semiconductors is theoretically possible.

Another potential cost saver is the manufacture of ac GPD is a reduction in the cost of etching the conductors on the glass panels. This has largely been done by hand, and typically 60 percent of the production has had to be scrapped. Automation techniques are expected to increase the yields from 40 percent to 70 percent. As costs decrease for the GPD, it is expected to attain a higher market share, especially in computer use.

DC plasma panels are popular as displays for applications requiring a few dozen to a few hundred characters. DC plasma panels are less expensive to drive than ac plasma panels on a small scale, but driving large scale panels becomes very expensive. Several companies manufacture dc panels of 480 characters (12 rows of 40 characters each). While larger dc panels of 960 characters are being developed, current line length is limited to 40 characters because of flicker problems in multiplexing more than 40 characters. This flicker problem does not occur with ac displays. Expanded dc displays will be attractive because dc plasma panels offer more color capability than ac panels (A.1, A.2, A.3).

Electroluminescent Devices ELD)

There are four basic types of electroluminescent panels: ac thick-film, ac thin-film, dc thin film cone and dc thick-film cone. While ac thick-film ELD represents the largest commercially produced device (principally used for lighting and indicator applications), ac thin-film technology, still in the laboratory stage, is the most promising for use in producing flat panels. At present, no one manufactures dc thin-film panels and only one company manufactures dc thick-film panels.

A typical thin-film electroluminescent panel consists of an electroluminescent layer (generally zinc sulphide doped with manganese, though other rare earth elements such as gallium arsenide diodes coated by erbium-activated phosphors can be used) sandwiched between two transparent insulating layers. This is further sandwiched between raw electrodes in the back and transparent column electrodes in front of a grid arrangement. When a high voltage of 150-200 volts is passed through the electrodes, the doped zinc sulphide material is excited, causing it to glow. The displays are limited to a single color, but a variety of different colors can be obtained by using different chemicals to form the electroluminescent layer. The primary color derived from the manganese is orange-yellow; other doping materials can allow red, green, blue and white light to be produced but at much lower display efficiency levels.

The basic difference between thick and thin-film panels is that thick-film technology employs powdery electroluminescent materials that are pressed into ceramic or plastic binders. The thin-film panels are made either by electron-beam sputtering, thermally evaporating or vacuum depositing the electroluminescent materials. While thick-film technology is substantially less expensive, it is not useful for alphanumeric and graphic displays. In addition, dc electroluminescent panels have shorter lifetimes than ac due to diffused impurities in the electroluminescent material.

The most promising approach is ac thin-film technology. Compared to other flat-panel technologies, ac thin film technology offers the potential of simple, low-cost, large flat panels capable of competing with the CRT. The ac thin-film ELD can operate over a range of -55° to +125°C, require little power, are very bright (typical luminous efficiency is 4 lumes/watt) and can withstand high altitudes and shocks. These devices offer the most potential, if and when four major programs are overcome: reducing the high voltages (150-200 volts) needed to drive ac ELD; reducing the cost of driver electronics for large panels; expanding the relatively short life (20,000 hours); and improving color efficience. (A3 & A4).

Arrays containing 12,000 elements, contrast ratios greater than 50:1, and power consumption below one watt have been demonstrated (10).

Liquid Crystal Displays (LCD)

This is the first example of a passive display, i.e., one that does not emit light but rather controls the passage of light through the display. The liquid crystal material possesses a phase, in addition to the normal solid and liquid-isotrophic phases, in which the material flows like a fluid but exhibits an anisotrophic crystalline state. Materials forming this nematic liquid crystalline phase from -10° to +80°C have been developed and are commercially available.

The liquid crystal material is usually contained between two closely-spaced (less than 25 microns) plates, both of which are coated on the inside with a conductive film. The optic axis is aligned by specially treating the electrode surfaces. The front conductive film is etched corresponding to the pattern to be displayed and is transparent. The back film is either transparent or opaque and reflective, depending on whether the LCD is to be operated in a reflecting or transmitting mode. In the latter case, a light source is mounted behind the display.

The electro-optical phenomena are of two types and occur when a voltage of a few volts is applied across the conductive films. In the first category, alignment of the LC molecules occurs via purely dielectric forces and therefore requires no appreciable current. Most research has been directed at these LC types due to their very low power and potential in electronic watch applications.

In one dielectric effect (the so-called 'deformation of aligned phases' or DAP effect), uniformly-aligned nematic crystals are deformed under the influence of an applied electric field, changing the optical transmission of the film. The second type of phenoma are the so-called dynamic scattering types, in which reorientation of the LC occurs via electrohydrodynamic effects and which require an electric current. These first-generation dynamic scattering LCD material is now rarely used, and the higher-performance twisted-nematic materials are now in favor.

There are several problems involved with producing large-sized LCD. The first is involved with the difficulty in reading the LCD from any viewing angle. Twisted-nematic displays require polarizers and the cost of polarizers increases exponentially with the display size. Newer dichroic materials, with color capability, point to the possibility of LCD competing in large panel applications, but these dichroic materials are substantially more expensive. Another problem occurs because the glass plates that sandwich the chemicals tends to warp when the area is expanded. There are many approaches to correcting these problems including the use of fiber optics adjacent displays and viewing screens.

Another major problem relates to humidity. More expensive glass-front-sealed LCD are immune to humidity. Lower cost plastic encased ones are not. Work is crrently being conducted on casing materials. Another problem concerns the slow response of LCD,

particularly at low temperatures. Newer LCD materials have been developed to allow practical operations between the range of -25°C to 90°C; however, most LCD materials work in a much smaller range (A.3, A.5, A.6).

Liquid crystal displays require only low voltages and very low currents in operation and have contrast which is relatively independent of ambient light level. However, viewing angle affects the perceived contrast and the response time is slow (100 msec), making drive of LCD matrices difficult. They are a strong contender for some character displays.

Electrochromic Displays (ECD)

Like LCDs, electrochromic displays are passive devices. As in other devices, ECDs employ paralled transparent plates on which appropriate electrode structures have been formed. The electrochromic material between the plates changes color under the application of a few volts between opposing electrodes and hence changes the color of the display. Contrast is excellent and wide viewing angles are permitted (20). Both solid and liquid EC materials are being investigated and results are very promising; however, this area is still relatively new and much work is still proprietary. Response times are slow, in the range of 40-100 msec.

Known research is being conducted on ECD using super-ionic conductors, a new class of ceramics developed initially for use in high-energy-density sodium-sulphur batteries. These hot-

pressured ceramics are used in batteries to separate the reactants while at the same time supporting a current flow between them by the motion of sodium ions in a sodium-aluminum substrate. The chronic compound used is tungsten trioxide. Significant problems still exist, including extreme temperature sensitivity.

Electrophoretic Displays (EPD)

These displays use still a different electro-optical material. Here, pigment particles of one color are suspended in liquid of another color. The pigment particles carry a charge, so that when exposed to an applied field, they move to the transparent surface electrode or away from the surface, depending on the polarity. Since the pigment scatters light when at the surface, it has the ability to change the perceived color of the display. Like ECDs, EPDs operate at relatively low voltages. They are also very slow and there is concern regarding the long-term reliability of the structure. These displays are at a relatively early stage of development. Proprietary research is being conducted by EPID (Electrophoretic Information Display), a newly formed venture started by Exxon.

Ferroelectric Ceramic Displays (FCED-PLZT)

These displays use a transparent ceramic material -- Ladoped lead zirconate-titanate. The electro-optical effect here is either electronically-controlled light scattering or electronically-controlled birefringence, depending on ceramic grain size. In both

cases, the applied electric field orients the ferroelectric domains, changing the optical properties of the material. Drive voltages here are higher (e.g., 40 volts) and the material must be excited transverse to the direction of transmission, causing fabrication difficulties. Poor contrast and stress-induced cracking are also concerns.

3.2.2 Technologies for Alphanumeric Character Displays

For alphanumeric displays, low operating voltage and compatibility with relatively standard IC technologies is important. High contrast is also important, especially under conditions of varying and sometimes high ambient lighting as in a cockpit. Light-emitting diodes and liquid crystal displays are certainly the most highly developed technologies at present for alphanumeric displays, and this is their main area of application. LCDs have advantages in the low currents required and in the relative insensitivity of their contrast to ambient lighting variations, although contrast is viewing-angle sensitive. LEDs produce a sharper, more attractive display at present and the development of high efficiency yellow displays should enhance their attractiveness for aircraft use. Both LEDs and LCDs are widely available now and will be widely used during the 1980s. Prices should decline slowly as manufacturing techniques improve. As for other display techniques, none appear to offer significant advantages over LEDs and LCDs for general alphanumeric character displays.

3.2.3 Imaging Displays

The dominant technology for imagi -type displays continues to be the CRT, which combines high resolution, large display area, high contrast, and high brightness with low cost. Color tubes are widely available and of excellent quality. Random-scan CRT terminals claiming 4096 x 4096 point resolution (which actually corresponds to the grain size of the phosphor) are available but shading is difficult and colors are limited. Raster-scan terminals offer 1024 x 1024 point resolution (4 times that of commercial TV) and an almost unlimited spectrum of color. Such imaging graphics terminals are still very expensive (several thousand dollars) but prices should come down as volumes go up and the hardware is more densely integrated. The number of support chips for CRT alphanumeric terminals dropped more than an order of magnitude between 1974 and 1979 (from about 150 to less than 20). Alphanumeric terminals are available today for less than $400 in small quantities.

The high beam voltage and relatively high ratio of tube depth to tube diameter are disadvantages of the CRT. Efforts at developing new flat panel displays are aimed at overcoming these two disadvantages but all appear to require significant tradeoffs in other areas, e.g., resolution or cost. The gas plasma displays are undoubtedly the most developed and appear the most promising. The fast response and sharp device threshold simplifies drive requirements and the display permits a wide viewing angle. Reliability and lifetime are potentially better than for CRTs. Cost is still significantly

above the CRT but should decline significantly as the technology matures. These displays have received substantial support from the TV industry. Display panels less than one inch thick have been demonstrated using GPDs. Drive voltage is reduced from several thousand volts for a CRT to less than 200 volts for the GPD; however, this is still too high for direct drive by most semiconductor technologies.

Many of the newer display technologies are not sufficiently mature to warrant comments on their potential for applications in communications terminals. Some are being developed in matrix form, but none appear to offer significant advantages over CRTs and GPDs for large-area imaging displays. For small area imaging in portable equipment and videophones, LCD matrices are being pursued, and 128 x 128 element displays having eight gray levels have been demonstrated (21). LED imaging displays containing 240 x 320 elements and 16 gray levels have also been developed (22).

CRT displays will certainly continue to dominate the image display area for the next several years. GPDs will possibly become practical for aircraft applications by 1985 or before, when reduced resolution may be an acceptable trade for reduced size and weight. Motivated by the entertainment market and by the proliferation of communications information terminals, the display area promises to be very active during the next decade.

The foregoing discussion has focused on the present approaches to two-dimensional displays. By the year 1990 most of these will either be in production or will have been dropped. Beyond this shakeout in technologies, it is entirely possible, and perhpas likely, that new display technologies will be discovered to supplement those considered above. Three-dimensional holographic displays may well be developed sufficiently by the year 2000 for some commercial applications (e.g., in complex industrial production and perhaps at entertainment (movie) centers) but it is not considered likely that these advances will be applied extensively in the home or by business within the time frame of this study. An area that probably will pervade society by 2020 and which will have a major impact in extending sophisticated man-machine communication capabilities to the minimally-trained user is that of speech recognition as described in the next section.

3.3 Man-Machine Communication by Speech

There are several reasons why speech communication between man and machine is of considerable interest. As outlined by Newell, et al. (30), perhaps the most compelling is the speed with which a human can speak (about four times faster than an accomplished typist), the fact that speech can be used by an untrained operator, is useful in a hands-free operational situation while freely moving about, and the possibility of data entry and immediate feedback without intermediate human transducers. In addition, in some situations, speech communication by computer offers considerable cost savings

through the replacment of salaried workers by machines (e.g. telephone operators). A secondary advantage is the possibility of verification of the speaker, increasing system security. For these reasons, man-machine communication by speech is the subject of considerable research in both universities and industrial laboratories.

There are several aspects to man-machine communication by speech. The first is voice response, wherein the machine syntehsizes a voice on the basis of stored information. The second is speaker recognition in which the object is for the machine to identify who the speaker is (speaker identification) or to verify that the speaker is who he says he is (speaker verification). The third and most difficult in speech recognition, in which the machine recognizes the words in a spoken message. Closely related to speech recognition (but with a higher level of difficulty) is speech understanding, in which the machine is prepared to take action based on the content of the message.

3.3.1 State of the Art

The areas of voice response and speaker identification are in a relatively advanced state of development while speech recognition is still in a relatively early stage of development (25).

Voice Response

There are three categories of voice response, depending on the type of information from which the speech is synthesized. The first category involves the storage of whole words in a quantized form. That is, individual words are spoken by a human and the

corresponding waveforms are quantized and simply stored in the machine to be recalled on demand (26). This method requires between 16,000 and 32,000 bits of storage for each second of speech to be synthesized. In the second category, the number of bits required to store the speech is greatly reduced by representing successive short segments of the speech by a fundamental set of parameters rather than storing the entire waveform (26). The set of parameters might typically be pitch, information as to whether the sound is voiced or unvoiced, and parameters which indirectly specify the position of the articulatory parameters of the vocal tact, such as frequencies of the first three formats (vocal tract resonances). This method requires between 600 and 9600 bits of storage for each second of speech to be synthesized, depending on the desired speech quality and hardware complexity. In the third category, the desired speech is simply stored as written text, and the machine performs the difficult task of inferring the requisite set of speech parameters from that written text as input to the second category of synthesizer (27). This method requires only about 75 bits of storage per second of synthesized speech, but of course, requires a much more involved procedure for synthesis.

For voice response, the two most important criteria of performance are the intelligibility and the naturalness of the synthesized speech. While intelligibility can be objectively measured by standard test procedures, naturalness is necessarily subjective and difficult to quantify. In many applications of voice response,

including most applications in which general public acceptance is not required, intelligibility is much more important than naturalness. All three categories of voice response synthesizers achieve a very high intelligibility. Thus, they can be distinguished primarily on the basis of naturalness.

For the first category of synthesizer, the individual words are perfectly natural sounding and intelligible, but the piecing together of the individual words to form the message is usually perceptually evident. The second category of synthesizer, due to the limitations of the model used to represent speech production, necessarily results in some loss of naturalness, particularly at the lower bit rates. There is the potential, however, of improving the transitions between words, since the machine has complete control over the pitch and other parameters and can match them at the word boundaries. In the third category of synthesizer, the speech is built up not from words, as in the first two categories, but rather from individual phonemes or syllables. Further, the sound intensity, sound duration, and voice pitch appropriate for each context in which the phoneme is inserted is different, necessitating a syntax analyzer program to examine the test and adjust these parameters appropriately. Shortcomings in these complicated algorithms result in some loss of naturalness in the synthesized speech, but this method has met a fair degree of success and should be considerably refined over the next few years (27).

The choice of a method of voice response is largely a matter of a tradeoff between the desired naturalness of the synthesized speech, the cost of the storage medium and the number of messages to be stored, the required processor time, and finally the cost of the hardware required to synthesize the final speech. Since large-scale integrated circuits to perform the latter are rapidly becoming available, the primary cost is the processor memory and time, the former being dominant for large numbers of stored messages.

An impressive example of the current capabilities in voice response is the Texas Instruments "Speak and Spell" educational toy for children. This device contains a custom integrated circuit for the voice response, has a vocabulary limited by memory size to 200 words, and sells for about $50.

Speaker Identification and Verification

Algorithms for speaker identification and verification have been very successful, at least when compared to the performance of human observers in the same tasks (28). The usual technique is for the machine to store reference information on a prearranged verification phrase for each of the speakers to be identified or verified. The person then speaks the verification phrase and the machine makes measurements on the offered voice sample and compares the results to the stored reference pattern to make a decision on the speaker identity. Typical reference patterns consist of temporal patterns of parameters similar to those used in speech synthesis, such as pitch, intensity, format frequencies, and sometimes other parameters

such as zero-crossing rates. In making comparisons of the actual speech with the reference, it is advantageous to use nonlinear time warping algorithms to account for the fact that a speaker will never say a word with exactly the same temporal pattern twice. Another important and not trivial problem is the automatic and accurate determination of the beginning and end of the utterance.

The performance of these algorithms is characterized by the error rate in speaker identification and the probability of both types of errors (false verification and false rejection) in speaker verification. Systems have demonstrated a 1% error rate in speaker verification in a laboratory environment and a 10% rate over a telephone circuit (25,29). When professional mimics were used, the laboratory error rate increased to about 4%. By comparison, under similar conditions, human observers had error rates about four times as great. These error rates are probably acceptable in some applications, such as access to bank account balances, but are not acceptable in applications where high security is required.

Speech Recognition and Understanding

With respect to speech recognition, the major distinction is between isolated word recognition and connected speech recognition. The isolated word recognition problem is very similar to the speaker verification problem, in that parameters are extracted from the speech waveform and compared to reference patterns, in this case one for each word in the recognition vocabulary, to determine which word was spoken. As in speaker verification, important problems are the deter-

mination of the beginning and end of the utterance, as well as time warping in the comparison with the reference pattern.

Successful systems for the recognition of isolated words can presently handle between 10 and 300 words in the vocabulary. Best performance is obtained when the reference patterns stored in the machine are speaker-specific, and are updated for a new speaker by having that speaker speak the words in the vocabulary into the machine (usually several times) and having the machine generate a new set of reference patterns. However, some systems have been demonstrated to work well for as many as ten speakers while using a single set of stored refrence patterns.

To a large degree the performance of isolated word recognition systems, which can be measured by error rate (where errors include rejections in which the machine cannot recognize any pattern) depends on the particular vocabulary. For example, the simple alphabet together with the ten digits include many sounds which are difficult to distinguish from one another, whereas a vocabulary of words chosen for some specific task are usually much easier to recognize because of fewer ambiguities. On the other hand, performance is largely independent of vocabularly size, unless the larger vocabulary includes certain words which are difficult to distinguish. The penalty paid for a larger vocabulary is a much larger processor time because of the larger number of reference patterns with which the utterance must be compared. Thus, some systems have been built

which recognize between 10 and 34 words and operate at or near real-time, whereas larger vocabularies inevitably result in slower than real-time operation for currently available processors and algorithms.

The current performance of isolated word recognition systems is an error rate of about two percent (30). Systems which operate at this error rate have been demonstrated for both microphone and telephone speech. This is an acceptable error rate for many applications, and in fact systems of this type costing in the neighborhood of $10,000 have achieved commercial acceptance (31). The main obstacle to wider application would appear to be not performance but cost.

The problems of recognizing connected speech are much more difficult. Because the number of sentences which can be formed with even a small vocabulary is prohibitively large, the machine cannot attempt to recognize whole sentences (which would be similar to isolated word recognition), but rather must attempt to break the message into individual components, and recognize them individually. Because words tend to be run together in connected speech, it is very difficult to reliably determine word boundaries. In addition, the acoustic characteristics of words in connected speech depend on context and as a result have a much greater variability in connected speech than in isolated words.

As a result of these difficulties, the reference pattern approach used in isolated word recognition is unsatisfactory.

Existing connected speech recognition systems therefore attempt to divide the entire utterance into accoustically invariant parts (called phonemes) and attempt to label each part as a phoneme or at least with selected features (such as whether it is a vowel or consonant). The machine then proceeds to a higher level of analysis, attempting to match these sequences of labels to a phonemic dictionary of the words in the vocabulary. Reliability and accuracy can also be dramatically improved by taking into account simple constraints imposed by context to assist in the choice of words. Thus, such systems work much better when not only the vocabulary is very small, but also the input utterances are related to a specific task for which the speech *understanding system* was designed.

Connected speech understanding systems have only been seriously pursued for the past five years. The existing systems typically have vocabularies of around 300 words, chosen for a specific task. They operate with a single speaker or at least require considerable training with each new speaker. The execution times of the programs is in the neighborhood of 20 to 30 times real-time. Word recognition accuracies of between 80% and 97% have been achieved, but of course the accuracy with which an entire sentence is recognized is considerably lower (since there are at least several opportunities to make a word error in a sentence), on the order of 50% to 80%.

The final type of speech recognition system which has received attention is the connected speech understanding system. For these systems, the criterion of performance is not the recognition of

each and every word in a message (as in a speech recognition system), but rather the intent of the entire message and response required in response to the message. Again, the system performs best when it is designed to be specific to some task, where the vocabulary is limited and there are more constraints on the messages which can be used to resolve ambiguities. The system has the advantage over a connected speech recognition system that first, the range of choices can be reduced, taking context and grammatical constraints into account; and second, individual words can be mistakenly identified and the overall message can still be understood correctly.

Several speech understanding systems are currently being developed, but it is premature to state performance figures for them because of the very early state of development. However, it is clear from the advantages stated earlier that their performance is potentially good, and better than that of the connected speech recognition systems. However, the full exploitation of these advantages is clearly a complex task, which will take a great deal of effort to complete and will result in systems which are slower (relative to real-time).

3.3.2 Predictions for the Future

Rapid progress in the performance of man-machine communication by speech systems is to be expected in the future. This is due in large part to the tremendous effort being expended in this direction, and in addition to the rapid increase in capability and decrease in cost of processors available for use in these systems,

and finally, to the expected availability of integrated circuits designed specifically for speech analysis and synthesis purposes.

Voice Response

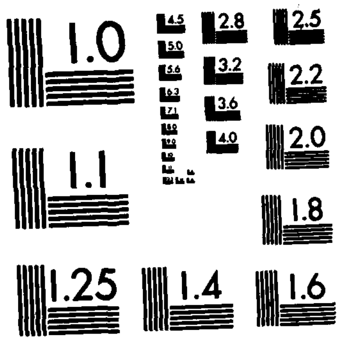
Progress is to be expected in the areas of intelligibility, naturalness, and the size of the vocabulary of voice response systems.

With respect to vocabulary size, this is largely a matter of the size of the memory. Particularly for synthesis of speech from text, the memory requirements are so modest as to not really represent a limitation even today. For the methods which require larger amounts of memory, the rapidly increasing size and decreasing cost of MOS, CCD, and magnetic bubble memories should result in impressive increases in vocabulary size. For example, assuming a doubling of memory size at equal cost every two years, devices of similar complexity and cost to the Texas Instruments "Speak and Spell" should have about a 13,000 word vocabulary in 1985 and a 400,000 word vocabulary in 1990. Since the average person uses a regular vocabulary of only about 15,000 words, we see that memory size will shortly cease to be a practical limitation on vocabulary size. Improvements in algorithms, resulting in reduced memory requirements for each word, will result in further advances in vocabulary size.

With respect to naturalness, a great deal of effort is being devoted to improving the naturalness of speech synthesized at a given bit rate, or equivalently, reducing the bit rate [illegible]

AD-A126 015 SOCIOECONOMIC IMPACT ASSESSMENT: COMMUNICATIONS INNUSTRY PHASE III TECHNO..(U) ACUMENICS RESEARCH AND TECHNOLOGY INC BETHESDA MD 02 FEB 79 FAA-APO-81-11-4 3/7

UNCLASSIFIED DOT-FA78WAI-932 F/G 17/2.1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

achieve a given level of naturalness. While a certain amount of progress is expected in this direction, it will probably not be dramatic. This is due to the fact that the lack of naturalness is due in large part to the inadequacies of models used for human speech production. Attempts at improving the models have demonstrated that the mechanisms of speech production are very complex indeed, and attempts at emulating these complex mechanisms have resulted in algorithms so complicated that execution times are very unreasonable (e.g., 10,000 times real-time). Thus, in practical applications, it is expected that, except for speech synthesis from waveform-encoded speech (requiring high bit rates) perfectly natural-sounding speech will probably not be available for some time.

The final criterion is intelligibility. Available voice response systems already achieve a high degree of intelligibility. In fact, intelligibility is probably not an impediment to any present voice response applications, and should improve gradually as efforts to improve naturalness also pay dividends in intelligibility.

Speaker Identification and Verification

Speaker identification and verification systems already outperform humans in these tasks. The question then arises as to how much better can they get? This goes back to the more fundamental question of the uniqueness of the human voice. Is each person's voice as unique as his or her fingerprint, or is it anything but unique, as, for example, hair color? The ability of mimics to fool a human observer suggests that a person's voice is not unique. On

the other hand, the inability of mimics to fool present speaker verification systems suggests an inherent uniqueness.

The situation would appear to be quite similar to that of the written signature. In that case, there are successful forgers (analagous to speech mimics), but detailed analysis of a forgery will reveal it as such. Thus, the written signature is a reliable enough verification technique to be accepted by the legal community, even though it may not be absolutely reliable. It is to be expected that speaker verification systems will achieve a similar level of reliability and acceptance within the next decade. If that is the case, then since the signature is accepted for access to a safe deposit box, so should a voice verification be accepted for access to a computer data file of all but the most sensitive nature.

It should be noted that viable alternatives to speaker verification in very secure database access applications are already becoming available, in the form of password and public key cryptographic techniques. However, these will require special purpose hardware and/or software at the access point, and hence speaker verification will remain attractive for the less demanding applications.

The manner in which speaker verification systems will be made more reliable will obviously be to incorporate more and more parameters in the reference patterns for each speaker, requiring a larger memory capacity per speaker and increased processing for comparison with the reference patterns. The algorithms for speaker

verification, rather than memory size or processor speed, place a practical limitation on speech verification accuracy in today's systems. This is because real-time operation is generally not required and only a single set of reference patterns must be compared.

Speech Recognition and Understanding

It is to be expected that future decades will see not only progress in the types of speech recognition and understanding systems discussed earlier, but also much more ambitious systems will be attempted and perhaps perfected.

With respect to isolated word recognition systems, there are likely to be dramatic improvements in vocabulary size and decreases in cost in the near future, primarily due to increased speed in processors and decreases in the cost of memory. Furthermore, it is to be expected that progress in algorithms for the more sophisticated speech recognition systems will pay dividends in the area of isolated word recognition, resulting in reduced computational loads for the same level of performance and the same vocabulary size. The bottom line should be systems available within a decade at a cost below $1000 for a vocabulary size in the 100 to 200 word range and real-time operation. Such a system would be suitable for most applications.

From the forecasts of Figures 2.11 and 2.15, we see an anticipated increase in memory size per chip which is much more dramatic than the processor cycle time reduction. Thus, the

latter is expected to limit the vocabulary size in the future. For an isolated-word recognition system based on a single processor, improvements in cycle time can be expected to result in an order of magnitude improvement in vocabulary size between 1980 and the year 2000, when vocabulary size should reach 200 or more. The addition of special-purpose hardware to do pattern matching in conjunction with a general purpose processor is expected to result in about a five-fold improvement in vocabulary size, with about 100 words available in 1980 and over 1000 achievable by the year 2000. These forecasts assume real-time operation, and it should be emphasized that relaxation of the real-time requirement or the addition of parallel processors will result in a corresponding increase in vocabulary size.

In the area of connected speech recognition and understanding, the problems limiting performance and accuracy are much more fundamental. The sources of knowledge which a human uses subconsciously in the recognition and understanding of speech are varied, sophisticated, and very difficult to reproduce algorithmically. They include (30) the characteristics of speech sounds (the easiest to reproduce), variability in pronunciations, stress and intonation patterns, sound patterns of words, grammatical structure of language, the meaning of words and sentences, and finally, the context of the conversation. The difficulty in incorporating all these elements

in a system is extraordinary, regardless of limitations in available processor speed. Thus, it is unlikely that these systems will approach the efficiency of the human in these particular tasks for some decades. Nevertheless, considerable improvement in performance is to be anticipated.

Connected speech recognition systems with a 200-word vocabulary cannot be reached by applying present-day algorithms to the faster processors forecast in Figure 2.11 (since a factor of 30 or 40 improvement in speed is needed). However, this goal should be obtainable at moderate cost by the year 1990 through a combination of algorithm improvement and special-purpose hardware for pattern matching.

Beyond these developments, the direction of reseach will be toward reduction in restrictions on vocabulary size, number of speakers, and speaking environment, and a broadening of the tasks for which the system is applicable.

The first goal of researchers, primarily due to the commercial possibilities, is a dictation system. Such a system would have a large vocabulary (say 10,000 words) and have the capability to recognize connected speech and generate typed text. A fairly high error rate, requiring considerable editing, would be tolerable since spoken speech is so much faster than typing. The primary impediment to such a system is the much slower than real-time

processing required for so large a vocabulary. This application requires, therefore, a major improvement in algorithms, and possibly, special purpose integrated circuitry to achieve real-time. The great effort devoted to this system because of its commercial potential should pay great dividends in other applications of speech recognition.

3.4 Other Components of Input-Output Devices

The control circuitry, codecs, displays, and speech recognition aspects of input-output devices have been discussed in previous sections. As indicated by Figure 3.1, however, there are many more components which could be associated with such terminals. While it is not possible to discuss them all in depth, this section will give comments on a few of the more important functions.

Among the remaining input devices, keyboards need little explanation. They are not expected to change significantly and are already highly standardized and relatively inexpensive. As the stored-program intelligence of terminals improves and, in the longer term, as speech recognition (and understanding) is added, it is likely that keyboards will be simplified to some degree.

The key device for video (picture) inputs is the imaging sensor itself. In the past, this function has been performed by a vidicon tube, which is relatively bulky and not extremely reliable. Solid-state imagers based on integrated-circuit technology have been in

development for about ten years. The basic transducer is a reversed-biased diode which collects the charge generated by impinging light. This charge is read out, and simultaneously the diode is recharged, once each frame time. CCD readout circuitry has become widely used during the last five years because of the lower clock noise associated with this technique as compared with earlier shift-register clocking schemes. Sensors have been designed as either line arrays or as full area arrays. In either case, the chips have been among the largest made by the semiconductor industry for any application, and this has resulted in very low yields and very high costs. For example, a 1,024-element line imager with imaging points spaced on 20 μm centers must be at least 20.5 mm (0.81 inches) long just for the array (ignoring any output circuitry). As a result, solid-state image sensors have found their primary applications in military equipment demanding high reliability under a variety of ambient operating conditions. In 1979, the state of the art in line imagers was 2,048 elements operating over a 500:1 dynamic range at a data rate of 500 KHz (32). Using deposited color filters, color-imaging area arrays having 484 x 384 elements and a 7.2 MHz horizontal clock rate (33) have been demonstrated. These area imagers are capable of imaging the frame size of Super-8 movie film and have exhibited excellent color quality. The development of such imagers is receiving considerable emphasis, particularly in Japan, since they are the key step in the realization of not only solid-state

video communications terminals but also video (movie) cameras storing images directly in the form of electrical signals on magnetic tape. This would allow permanent (non-degrading) storage on an erasable, reuseable medium requiring no development cycle and therefore allowing virtually instant viewing over a television display screen. The development of such 'instant replay' systems for home use, either in communications terminals or as stand-alone video cameras will likely occur during the 1990s, the only practical barrier being cost. To overcome the cost barrier, the application of redundancy to the sensing chip is essential. While it is likely that redundant elements will not be included in the imaging array itself, it should be possible to identify and tag bad array elements at manufacture and correct these pixels as they are read out -- probably by assigning them a value dependent on those of their neighbors. While this procedure cannot correct bad pixels, it can minimize their visual effects, making arrays with a considerable number (up to one percent) of bad elements useable and reducing chip costs between one and two orders of magnitude. Large area imagers cost from one to two thousand dollars each today in perfect or nearly perfect operating condition.

Both magnetic tape and disk systems are expected to improve substantially during the next twenty years. Magnetic tapes, cassettes, and cartridges offer the lowest cost per megabyte of storage at present (less than 40 cents) and are useful in situations where a high-capacity serial bulk store is needed and fast access is not.

Cartridges measuring 4 x 6 inches and containing more than 75 M bytes of formatted, error-corrected information have been announced with average transfer rates of 20 K bytes/second and burst transfer rates of up to 4 M bytes/second. Storage densities are of the order of 10,000 bits per inch (34, 35). Figure 3.5 compares this cartridge system (the 3 M HCD-75) with other tape and disk systems. The present interfaced cost of the system with a single drive is less than $2500. In the future, systems with more than an order of magnitude more storage capacity, increased speed, and lower cost should evolve. Such systems should find important use in a variety of terminal applications requiring high-volume bulk storage.

For applications requiring relatively fast access, high capacity, and moderate cost, a variety of disk approaches are possible (36). Low-cost floppy disks offer 1.6 M bytes of storage capacity and a total interfaced cost of well under $1000. Off-line storage density is extremely high. Higher capacity and faster performance is available from 8-inch hard disks. The associated disk drives now sell for between $2000 and $4000, offering from 4 M bytes to over 50 M bytes of storage, average access times between 30 and 60 msec, and data transfer rates of 0.5 to 1.2 M bytes/second (37). These systems, which use IBM Winchester (sealed, fixed head) drive designs, are relatively new, and their use is expected to grow from 4100 spindles in 1979 to more than 60,000 in 1982. As they mature, performance should improve and prices should fall.

A Comparison of 1980 Tape and Disk Drives in Terms of Cost Versus Capacity (34)

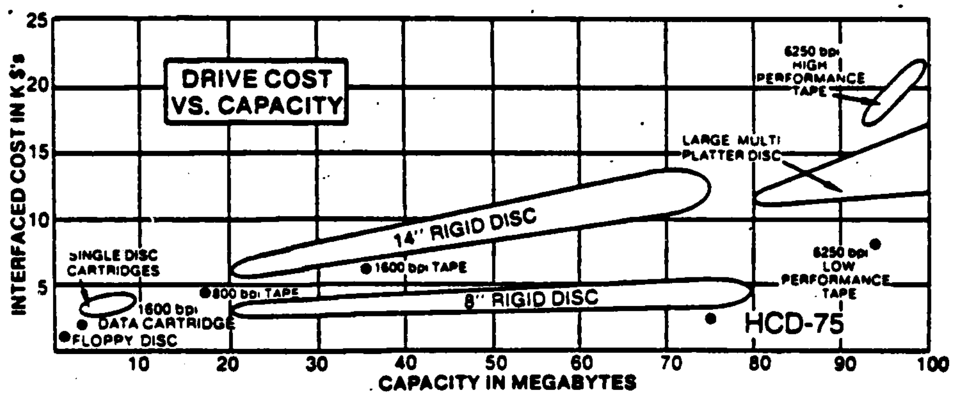


Figure 3.5

For lower capacity but smaller size and faster access, magnetic bubbles offer a monolithic non-volatile storage medium free from moving parts as described in Section 2. Overall, there exists a nearly continuous spectrum of memory technologies, increasing in capacity and decreasing in cost per bit and speed and extending from semi-conductor RAM to bubbles, disks, and tape cartridges. Each technology fills a need, and it is likely that each will play an important role in future telecommunications systems. Newer approaches such as electron-beam addressable memories, optical, and holographic stores will likely have to compete with and displace one of these established approaches if they are to make a significant contribution in future systems.

A final function needed for communications equipment is fulfilled by the printer, whose hard copy output represents the most convenient data form for many applications. Present printers can be generally classed in two groups -- line printers and serial printers. Line printers print a line at a time at very high speeds and are used primarily in shared, centralized facilities where their high throughput is useful. Page printers (e.g., the Xerox 9700) are included in this class. Serial printers, which print a character at a time, are lower in throughput and cost and are more likely to be found in small dedicated applications in business or the home. Typewriters would be included in this class.

Line printers can be subdivided into impact and non-impact types, with impact mechanisms accounting for about 87 percent of the present market (38). Impact printers use drum, chain/train, band/belt, and distributed-impact-matrix printing techniques while non-impact devices are based on electrostatic and Xerographic techniques.

Drum impact printers were among the early line printing approaches and use a drum with characters embossed around its periphery. When the characters have rotated into the proper position opposite the paper, a set of hammers force the paper and ribbon against the paper to print the line.

Chain or train printers use a series of character slugs which are either connected or unconnected and which are either pulled or pushed around a horizontal track. Several complete character sets are included in each chain/train which move past hammers (one per column) at constant speed. As the appropriate characters appear in position, the hammers print them to form a line.

In band/belt printers, the characters are etched on the periphery of a steel belt or band. Like the drum and chain/train printers, a set of hammers is used (one per column) to print the page as the characters rotate into position. These printers offer somewhat higher speed (up to 3800 lines per minute as compared with 2000 lpm for the other approaches), interchangeable fonts, and good print quality.

Distributed-matrix printers employ a set of impact wires which oscillate horizontally opposite the page. As the paper moves past the wires they can be actuated to impact the paper, forming a series of dots. As the paper moves through seven steps, a complete 5 x 7 dot character matrix is formed. Impact-matrix printers are lower in both speed (600 lpm) and cost. Overall, all of the different types of impact printers can print multipart (carbon copy) forms and are relatively reliable, but they are also noisey, a disadvantage overcome by non-impact printers. Non-impact printers employ dot-matrix-generated type fonts with up to 300 x 300 dots per square inch. These printers use either plain paper (xerographic or electrophotographic printing) or treated dielectric paper (electrographic printing). In xerographic printing, the most widely-used approach modulates a laser which is directed onto a line segment of a drum photoconductor moves past the line segment to form the image. Toner is then attracted to the photoconductor over the image areas, producing a toner image which is transferred to the paper and fused into a permanent image. These printers are versatile, very high in print quality, high in throughput, and high in price.

Electrostatic (electrographic) printers pass specially-treated paper over an array of fine metal style (37). Each stylus is charged or not depending on the image to be printed, and the charge is then transferred to the paper as a series of charged spots. The paper is then passed through a toner, where the charged spots attract ink particles. These printers are lower in price (and quality) than the xerographic printers and require special paper.

Figure 3.6 compares the existing impact line printers in terms of cost and throughput. Cost ranges from under $2000 for one impact-matrix device to over $100,000 for a band/belt printer, with respective printing rates of from 120 lpm to nearly 4000 lpm. Figure 3.7 compares the 1979 and 1985 printing markets. In spite of an overall increase from $900 million to $1.6 billion in the printer market, the price of an average unit is expected to decrease by about 30 percent during this period. Band/belt and impact matrix printers will dominate the 1985 market, accounting for 52 percent of the total, while non-impact printers will increase their market share from 13 percent (1979) to 41 percent (1985). Impact printers are expected to èxceed 4000 lpm by 1985, while non-impact printers should exceed 200 pages per minute (the Xerox 9700 now prints at 120 ppm). By 1990, more than half of the line printer market should be filled by non-impact machines. Throughputs should rise and prices should fall at perhaps 10 percent per year throughout the remainder of this century. Most market growth is expected in two areas: high quality printers capable of handling color photographs as well as text, and and low-cost printers for small business and home applications. In the latter area, some overlap with serial printers can be expected.

A Comparison of Impact Line Printers -- Throughput Versus Cost (38)

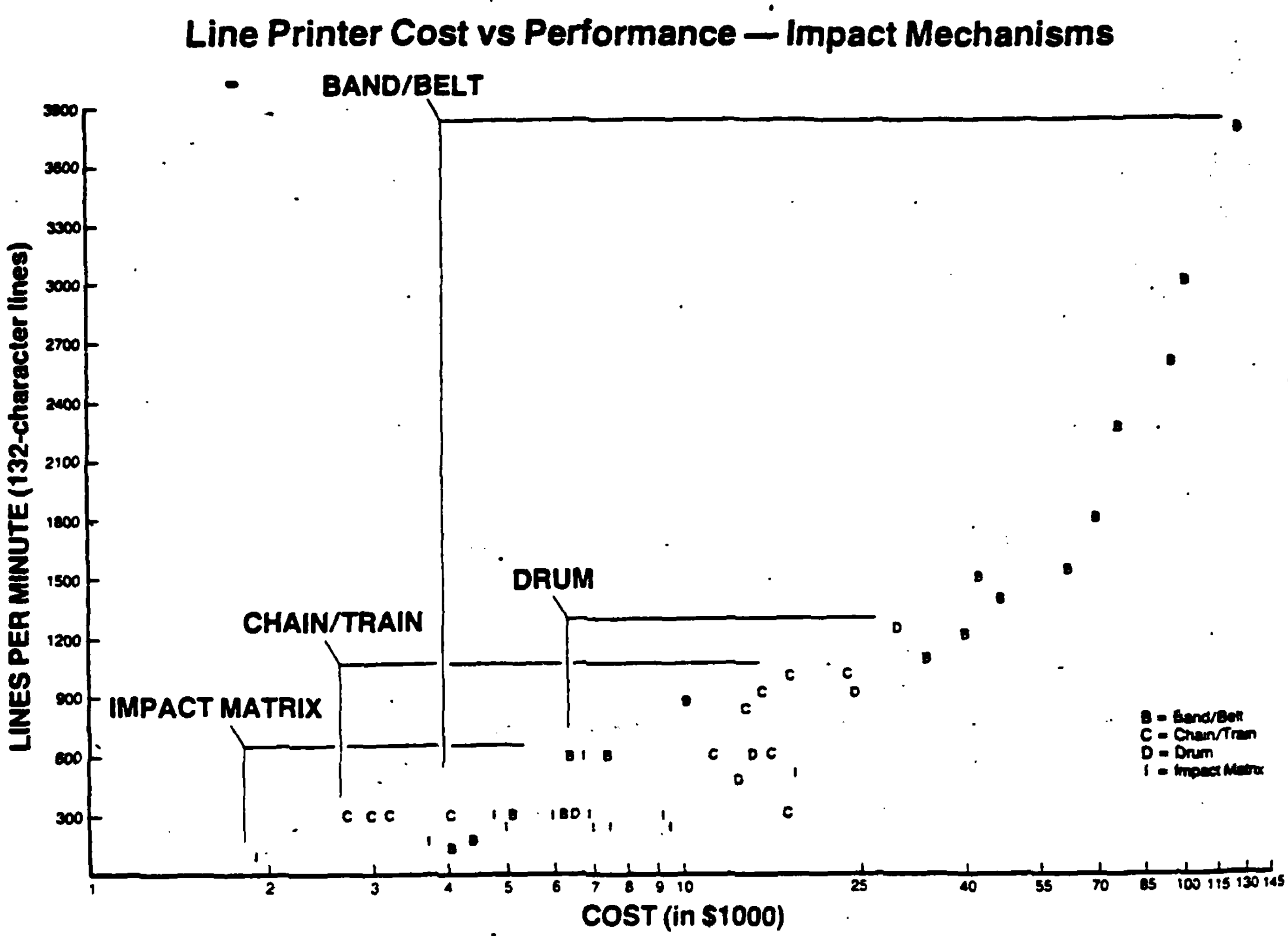
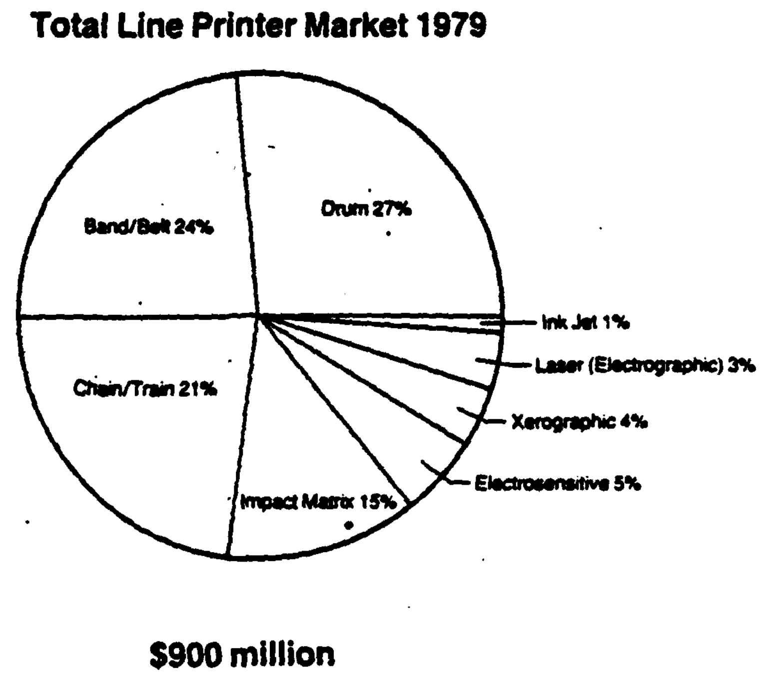


Figure 3.6

ACUMENICS

Line Printer Markets, 1979 and 1985 (38)



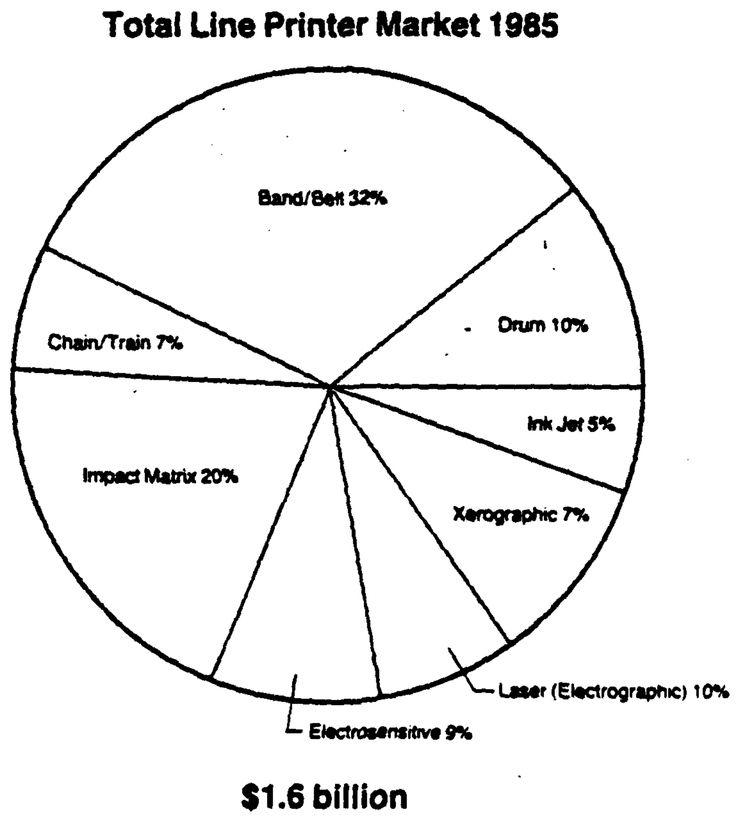
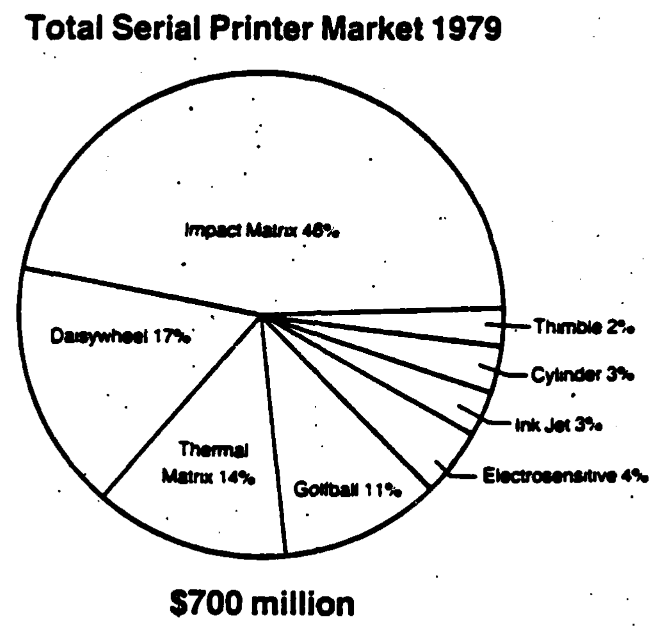


Figure 3.7

A substantial growth in the market for line printers can be expected during the early years of the next century if library functions become linked as part of a nationwide (or worldwide) telecommunications network.

Serial printers divide into impact and non-impact types and print characters in a one-at-a-time serial fashion. At present, nearly 80 percent of the serial market is held by impact printers, which use a variety of mechanisms (impact matrix, daisywheel, golfball (Selectric), thimble, and cylinder) (40). Impact matrix printers hold the largest market share, as shown by Figure 3.8. Fully-formed character printers, which presently have the best print quality, are expected to decline as a percentage of the total market, yielding to the impact-matrix approach and to non-impact approaches. The value of the printer market is expected to more than double (in present dollars) by 1985 while the number of printers shipped will nearly triple. Word processing systems for business applications are expected to be the most dramatic growth area. At present, serial printers range between one and six thousand dollars in cost and print from 10 to 600 characters per second. The average cost is about $2100, which should drop to $1800 or less by 1985. With the proliferation of computer-based equipment and the expanded capability of the telecommunications network, serial printers should exhibit their strongest growth between 1990 and the year 2000. Most of this growth is expected to occur in non-impact designs.

Serial Printer Markets, 1979 and 1985 (40)



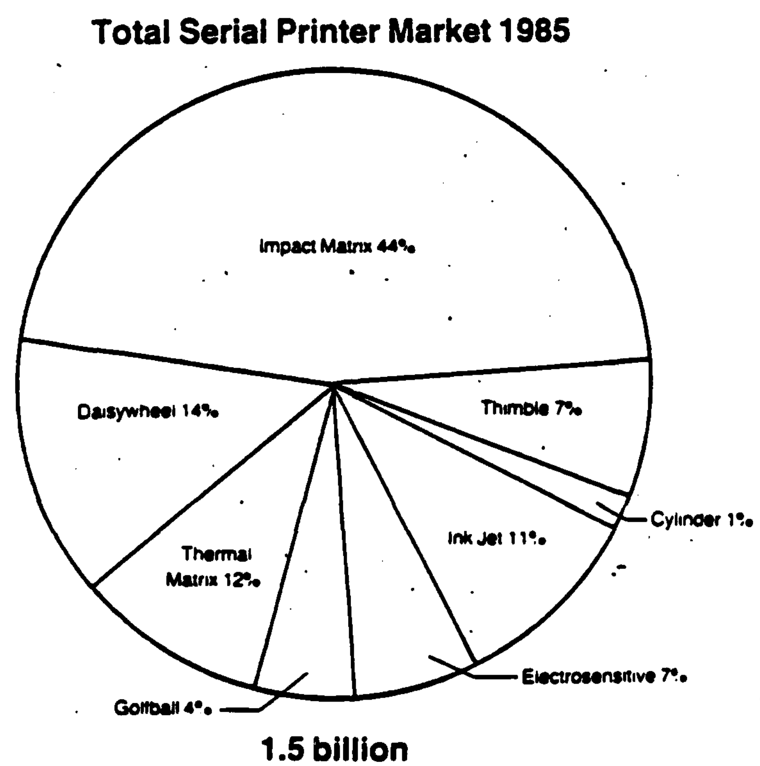


Figure 3.8

ACUMENICS

Non-impact serial printers are quiet, have fewer moving parts, and are potentially lower in cost and higher in throughput than their rival impact printers. Present non-impact devices include thermal, electrosensitive, and ink-jet types. Thermal printers transfer the desired character to the paper via heat, which selectively darkens the thermally-sensitized paper. Individual wires in the print head form a dot matrix. Electrosensitive printers also rely on special paper and work by eroding a thin layer of aluminum from the paper surface to expose a darker base.

Ink-jet printers use a matrix of tiny nozzles to blow ink on the page as needed to form characters. Both drop-on-demand (pulsed) and continuous-flow systems are being developed. In the continuous-flow systems, electrostratic deflection is used to direct the ink onto the paper or not, depending on the character. Ink-jet printers are still relatively new with continous flow machines such as the IBM 6640 printing at 92 cps and costing as much as $24,000. This approach is regarded as very promising by many, including IBM, and could capture a major share of the market by the 1990s.

It would be naive and short sighted to think that all of the approaches to line and serial printing have been identified as of 1980. It is in fact likely that during the time span of this forecast several new approaches will be tried and perhaps adopted, replacing the techniques discussed above. The present approaches should therefore be regarded only as a base on whose technology the

printers of the year 2000 will improve. Serial printers are expected to continue to occupy the $500 to $5000 price range, although improvements in quality and throughput can be expected. Substantial overlap between high-end serial printers and low-end line printers is likely.

3.5 An Unconstrained Technology Forecast of Terminal Functions

In past sections, most of the major components associated with future input-output terminals were discussed relative to their operation and near-term future. In this section, descriptive parameters associated with these components will be summarized and expressed as quantitative unconstrained forecasts. These forecasts will cover the time period from 1980 to the year 2000. Over this time span we will be primarily concerned about the development and evolution of technology and not with its diffusion into society, which could take considerably longer. In subsequently expressing the constrained forecasts, where diffusion will play a more prominent role, the time horizon is extended to the year 2020.

By the year 2000, it is expected that the present technologies will have matured and radically different approaches will be required to make substantial further progress. While such approaches will undoubtedly come, it would be futile to try to define them here.

Given the relatively long diffusion times associated with telecommunications, approaches developed after 2000 are not likely to be widely deployed by 2020 anyway, and the technology of 2000 is likely to be sufficiently advanced to support most presently-conceived applications.

For telecommunications terminals, a number of parameters might be forecast, including speed, power, size, weight, cost, and reliability. However, the diversity of the components involved make forecasts of some parameters difficult. It is also doubtful that some parameters will change significantly or would be of much significance for terminals even if they were forecast. For example, the basic telephone, which will continue in largely its present form, is unlikely to change significantly in any of the above categories except possibly power, and this change will be more meaningful for its effect on the central office (switching) than for its significance to the terminal function itself.

The reliability of input-output devices can be expected to improve substantially in some areas as the result of advances in VLSI control circuitry, reduced interconnections, and the simplification of many mechanical functions. This is particularly true for some of the relatively less reliable present components (bulk stores and printers). While the conversion away from moving parts will not be total, system reliability will nonetheless benefit substantially from increasingly electronic control. The less reliable of present components should improve by one to two orders of magnitude in reliability, while components such as the telephone,

which already offer extremely low failure rates, will remain relatively unchanged.

Weight in input-output devices is expected to generally mirror system size as it did for the systems considered in Section 2.7. It is not expected to be of major significance except for airborne applications where it will depend on the specific system implemented. Specific communication systems for NAS applications are considered in Phase 4 of this study.

Speed and power are interrelated for electronic logic functions but not for displays, bulk storage devices, or printers. The power dissipated by terminal electronics is not expected to be a major factor in the overall terminal requirements, which will be dominated by displays, bulk stores, and printers as was the case in the instrumentation system of Section 2.7. These components will be locally powered, with telephone power likely to remain in the central office. Speed is task-oriented, and must be defined for the particular component or function needed. Component speeds will be considered here, while specific system considerations are left to Phase 4.

In order to summarize some of the more important aspects of terminal components, we will define a terminal consisting of a number of elements and forecast the size and cost of that terminal over the 1980-2000 time period. Since power and weight are not expected to be significant factors driving the development of such

equipment, they are not forecast separately. Comments and forecasts of system spêed are given in conjunction with specific component discussions.

The terminal system to be considered is assumed to consist of the following parts:

1) Microcomputer Control Complex

This portion of the terminal consists of a VLSI microprocessor having a complexity of at least 10,000 gates and an instrution cycle time of 100 nsec or less. The processor is supported by 320 K-bits of ROM, 1 M-bit of high-speed RAM, and 16 M-bits of magnetic bubble memory. In addition a variety of peripheral interface chips are included.

2) Bulk Storage Module

The terminal is equipped with a bulk storage device, taken here as a tape cartridge system, having a single-drive single-cartridge capacity of 75 m-bytes (600 M-bits) and an average transfer rate of at least 200 Kb/sec.

3) Hard Copy Output Device

The terminal is provided with a printer capable of character or graphic reproduction. The printer has a throughput of several lines per second. Normally driven through an electronic data link, it is also accessible via system keyboard.

4) Video Input System

The system is assumed equipped with a solid-state imager capable of real-time transmission of TV quality pictures and graphic data.

5) Video Display

The dominant form of display for the terminal is a 25-inch TV display screen, equipped for showing commercial broadcast TV, Teletext, or wired data.

6) Miscellaneous Input-Output Devices

The terminal has a standard microphone and speakers for audio I/O, telephone/videophone transmission. It is also equipped with a keyboard for data entry.

This terminal is intended to represent a broadly-equipped general-purpose input-output device. Equipped as described above, it could provide the following functions:

1) Broadcast or wired TV.
2) Teletext and viewdata access to news and library facilities.
3) Electronic mail.
4) Building security and environmental control.
5) Telephone and videophone service.
6) Electronic funds transfer.
7) Remote shopping.
8) Electronic games and learning.
9) Broadcast music.
10) Computer-based problem solving and record keeping.

Figure 3.9 shows the relationship of these various functions to the different components of the terminal system. Most functions depend on the microcomputer and at least a portion of its memory. This memory is extensive enough to permit the recognition of human speech for security, control, and perhaps dictation purposes and would likely allow an input-output vocabulary of several hundred words. The software and dedicated hardware necessary for this function should be available by 1990. While most of the applications are oriented toward the home, the terminal would be equally useful for small (and large) businesses. It would also be useful for a variety of administrative, maintenance, and ATC functions within the FAA.

Figures 3.10, 3.11, and 3.12 forecast the cost and size of the terminal from 1980 to the year 2000. Cost is expected to drop by nearly an order of magnitude during this period. The most dramatic decreases are expected in solid-state memory, which accounts for almost 50 percent of the total system cost in 1980 and only 6 percent in the year 2000. Peripherals (display, printer, tape drive) increasingly dominate the system. Size is expected to decline by a factor of only about three, and this factor is very sensitive to what happens in the display area. The microcomputer, including the bubble memory, is expected to fit on a single printed circuit card in the year 2000, occupying well under ten percent of the system volume.

The microprocessor chosen with this system will probably not be available in 1980 at the sub-100 nsec cycle time stated, although it should be available before 1985. By 1990 such a processor should be comfortably behind the state of the art; nevertheless, it is expected to be capable of managing most terminal functions very well.

The read-only memory (320 K-bits) will hold many of the fixed routines needed by the terminal. The high-speed ROM and RAM are implemented (Figure 3.10) with chips close to the state of the art before 1990 and are therefore somewhat more expensive than smaller chips. It is important to recognize that these costs assume high volume production and are close to production cost. They, therefore

Relationship of Terminal Functions
to Terminal Hardware

	Microcomputer	Bulk Memory	Printer	Video Input	Video Output	Misc. Peripherals
Television	X				X	
Library Access	X	X	X		X	X
Mail	X		X	X		X
Security/Control	X	?				X
Videophone	X			X	X	
Funds Transfer	X		?		X	X
Shopping	X		?		X	X
Games/Learning	X	X			X	X
Music						X
Computer Calc.	X	?	?		X	X

Figure 3.9

Terminal Cost and Size Forecast by Year, 1980 - 2000

	1980	1985	1990	1995	2000
Microcomputer					
Processor	160	60	20	15	10
ROM	256	64	32	25	20
RAM	1000	400	100	50	35
Bubble	8000	3200	800	250	100
Peripheral LSI	400	200	120	80	60
	$ 9816	3924	1072	420	225
Chassis/ Boards	175	125	60	60	60
A,W,T	85	65	28	26	26
Total Cost	$10076	4114	1160	506	311
LSI Chip Count	(107)	(43)	(18)	(14)	(12)
Bulk Memory	2500	1800	1000	800	700
Printer	3000	2200	1500	1000	800
Video Input	2750	1150	350	175	150
Video Display	600	500	400	350	300
Misc. Peripherals	700	550	425	315	230
Total Terminal	$19626	10314	4835	3146	2491

SIZE (cubic inches)

	1980	1985	1990	1995	2000
Microcomputer	1000	600	200	160	140
Bulk Memory	1500	1500	1200	1200	1000
Printer	5000	4500	4000	3500	3000
Video Input	200	150	100	100	100
Video Display	30000	25000	21000	8000	8000
Misc. Peripherals	2000	1800	1500	1500	1500
Total size	39700	33550	28000	14460	13740

Figure 3.10

Terminal Cost Forecast, 1980 - 2000

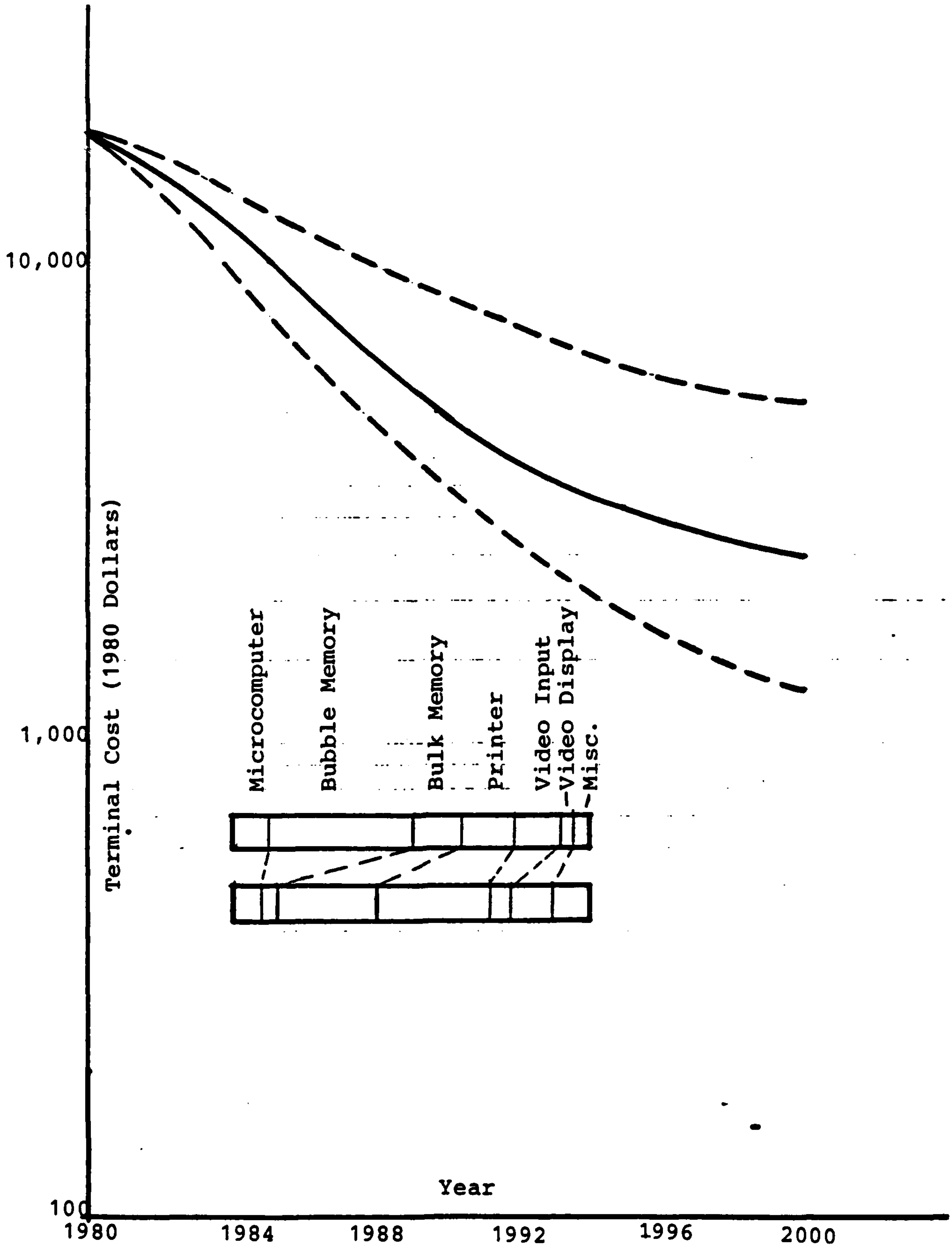


Figure 3.11

Terminal Size Forecast, 1980 - 2000

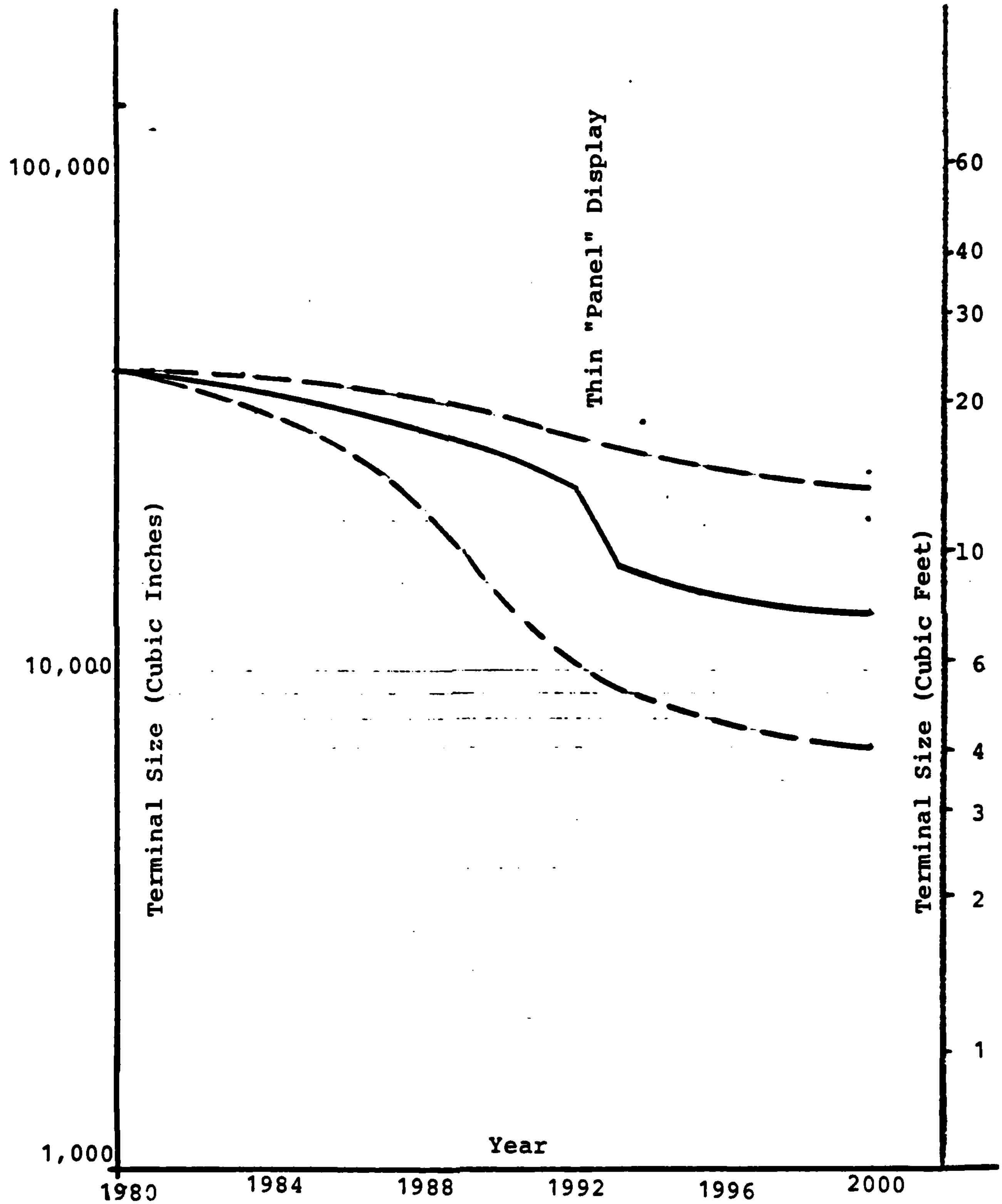


Figure 3.12

imply a highly competitive market in which supply at least equals demand, in spite of a rapidly escalating demand. A seller's market, in which demand far outpaces supply, could seriously retard the decline expected in memory cost. Most program execution is expected to take place from RAM.

The bubble memory is expected to act as a relatively fast bulk store into which specific algorithms, routines, and other data can be paged from tape. Bubble memory entered the commercial marketplace in 1979 at the one megabit/chip level after nearly a decade of development and some disappointing earlier ventures at the 92 K-bit/chip level. The megabit chips look significantly better, offering a 200 KHz data rate and an access time of 20 msec. However, the interfaced cost per bit is still relatively high and as a practical consideration, the 16 M-bit memory would probably be implemented as a disk in 1980 at somewhat lower cost and slower speed. By 1985 the bubble memories should be sufficiently advanced as to constitute a preferred (or at least competitive) choice for a fast access 16 M-bit store. Chip capacities of at least 4 M-bits should then be available and access times below 10 msec are expected. By the year 2000, 16 M-bit bubble chips with access under 0.1 msec are forecast (41). Nevertheless, the use of a bubble store in 1980 probably results in a slightly higher initial cost for the terminal than would a disk. Data transfer rates on bubble memories should exceed 1 MHz in the near future and will likely reach 10 MHz by the 1990s.

The microcomputer complex is assumed to require a substantial number of LSI support chips to handle peripheral functions, including one or more codecs and chips to control information protocols. As time passes, some of these jobs will be taken over by the microprocessor, while other functions will be combined on a single chip. General forecasts of power, speed, and cost for VLSI logic are given in Section 2.

A tape cartridge system is assumed for a high capacity memory store for the terminal. The added access delays compared with a disk are assumed tolerable since the bubble memory will provide a higher speed cache to RAM and effectively replace what might be an underutilized disk. It is expected that small business and other applications will create enough pressure for low-cost peripherals that the cost of a fixed function will decline by a factor of between three and four for tape drives, small disks, and printers during the next twenty years. This amounts to a yearly decrease of only six percent, which probably implies a price rise in present dollars. The potential exists to do substantially better, accounting for some of the uncertainty in Figure 3.11. The sizes of tape drives, disks, and printers are not expected to change significantly. The printer assumed for the system would be most likely a low-end line printer, either impact matrix or non-impact in the near term. Longer term, some form of ink jet or xerographic device might be employed.

The video input device is assumed to consist of a solid-state imager chip, lens assembly, and peripheral data processing chip. It is assumed capable of 512 x 512 resolution for real time TV-quality imaging of scenes or graphic data. In 1980 this system is not available, although the price is representative of present area imagers of somewhat lower resolution. The rapid drop in cost reflects anticipated demand and the application of defective element correction in the signal processing chip. By 1990, the imager should permit high quality color transmission. The video display reflects a commercial broadcast color TV with provision for teletext reception. The microcomputer and its memory would also participate in teletext recording and display. Among the miscellaneous items required for the terminal would be a keyboard, audio pickup device, AM/FM receiver, and speakers.

Examining the cost of the terminal as shown in Figure 3.11, its 1980 cost of $20,000, which reflects only the manufacturing cost of the computer, is too high for home use but certainly within range of most small businesses. Businesses are indeed starting to purchase systems with many of the features of this terminal. By 1990, however, the system cost is expected to drop below $5,000, so that during that decade penetration into the home market should begin. Certainly by the year 2000, terminals such as the one described here will be practical for home use. Thus, terminal capabilities and cost are not expected to slow telecommunication developments and could increase the pressure on switching and

transmission devices for new services and features. It should be noted that the size of the entire terminal is roughly the size of a present color TV, and that much of the uncertainty in system size reflected in Figure 3.12 is generated by uncertainty as to what display will be used. It is assumed that in the early 1990s a significantly thinner pseudo-panel color display will be available, whether it is a form of CRT or some other device.

3.6 Filtering in Telecommunication Systems

Before leaving the subject of terminals to examine the switching and transmission functions, comments should be given relative to two additional topics: filtering, and the subscriber line interface. Filtering is treated in this section, and line interfaces are discussed in Section 3.7.

Filtering is an essential part of any communication system where signal sampling and multiplexing are employed. The typical 8 KHz sampling rate in telephone systems implies the need to bandlimit individual signals at 4 KHz to prevent foldover or aliasing distortion between adjacent channels. Hence, low-pass filters having sharp cutoffs near 4 KHz are required. Additional filters are required to aid in decoding dual-tone multifrequency (DTMF) signals such as TOUCH-TONE, which use the same spectrum as speech. Filters are a significant cost component of communication systems, whether physically placed in the transmission (repeater), switching (line interface, digit receiver, channel bank), or terminal area. As digital signaling is increasingly used and more complexity is placed

in the terminal, filtering will become more closely associated with the input-output area.

The earliest filters employed passive LC circuits. These were bulky, heavy, and expensive, particularly for the audio range needed for telephone communications. Transistor circuits permitted the development of active RC filters, and with the development of low-cost monolithic operational amplifiers, these active filters were improved significantly in terms of cost, size, and selectivity. Since integrated resistors and capacitors are limited in the range and precision of allowed values, these passive elements have frequently remained off-chip, being realized as thin film or discrete elements. Thus, the size and cost of recent active RC filters have tended to be dominated by the discrete capacitors.

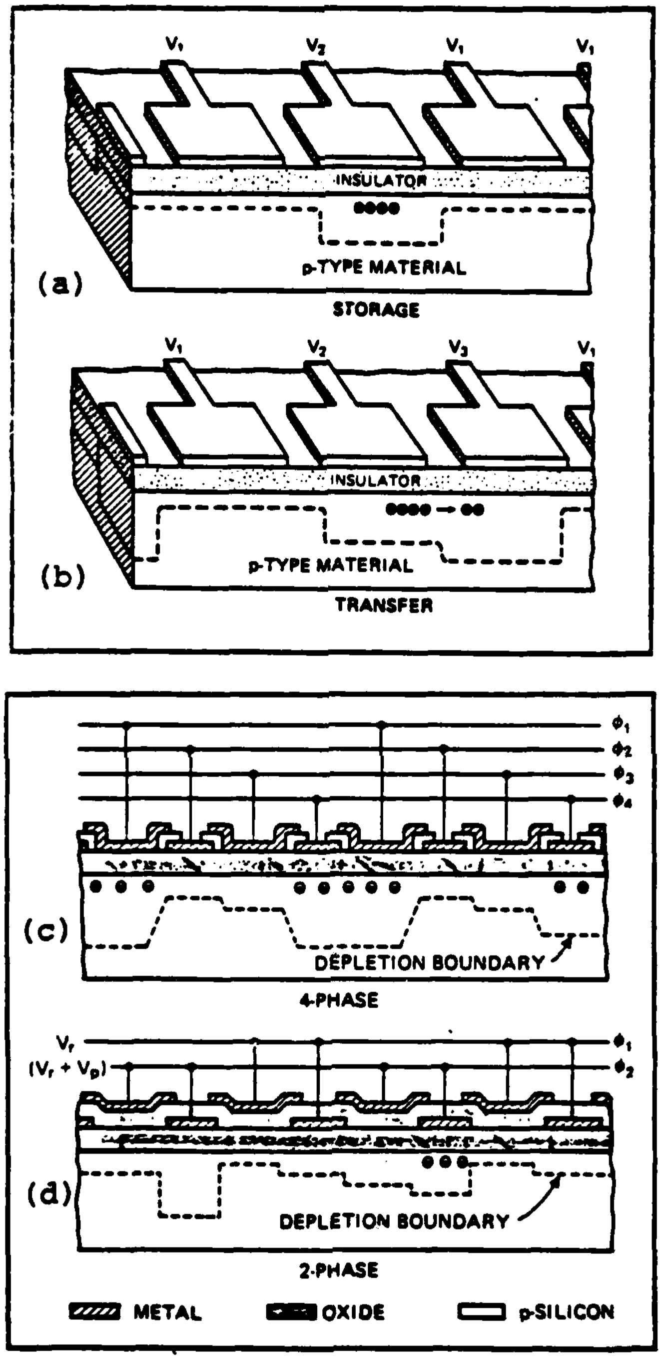
There are three approaches which offer alternatives to active RC filters which offer the prospect of further cost reduction and miniaturization: digital filters, CCD filters, and switched-capacitor filters. Digital filters require an analog-to-digital converter to digitize the analog signal and implement the filter characteristic by means of a computer algorithm. This approach has advantages in terms of flexibility and real-time programmability, but in communication systems, where most filter characteristics are well defined and not subject to change, these advantages are relatively less important. Digital filters require a high-speed processor and relatively large amounts of memory, and in spite of

recent progress in digital integrated circuits, such filters are still not practical for telecommunications. Even in the late 1980s, when processor speed will be compatible with filter requirements, it is likely that the analog CCD and switched-capacitor techniques will be preferred.

Charge-coupled devices (CCDs), acting as analog delay lines, offer one practical means for implementing monolithic filters which are stable and free from discrete components. CCDs consist of a series of parallel electrodes on an oxidized silicon substrate as shown in Figure 3.13 (42). The input signal is sampled and a charge proportional to the input signal amplitude is injected into a depletion area (potential well) under the first electrode. By alternately pulsing the electrodes from two, three, or four-phase clocks, the charge packet can be moved along the electrode array, finally emerging at the end of the array where the input signal is reconstructed, delayed by a time equal to the clock frequency times the number of stages.

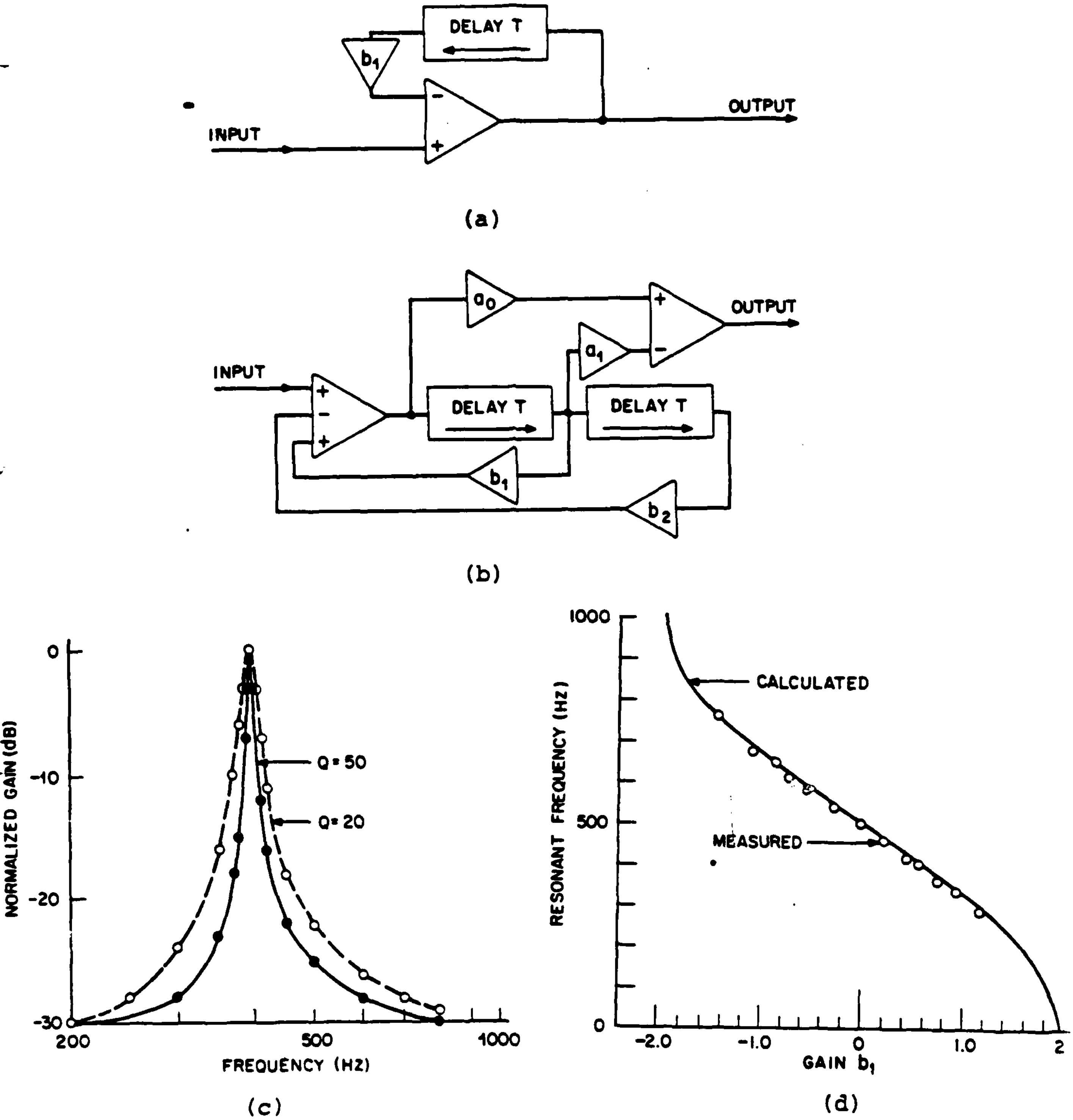
CCD filters can be of the recursive or the transversal types. Recursive filters are shown in Figure 3.14 (43), where feed back is used to form a resonant circuit. The center frequency is set by the feedback coefficients and by the delay, which in turn is a function of the clock period and the number of stages. Since crystal clocks are extremely stable, so is the filter characteristic. Transversal filters are realized as the weighted sum of many samples of an input signal, each taken at a different time. The samples are realized as outputs of the various stages of a CCD, and the weights are defined

Charge Transfer in CCDs (42)



(a) In a three-phase charge transfer, charge is stored in a potential well formed by a voltage V_2 larger than V_1; (b) Charge transfer is accomplished by applying a voltage V_3 greater than V_2, causing charge to spill over; (c) four phase and (d) two phase clocking schemes using two levels of metal.

Figure 3.13



(a) First-order recursive filter, and
(b) Second-order recursive filter with adjustable coefficients and hence a tunable bandpass
(c) Amplitude response for a 10-element CCD/BBD similar to (b)
(d) Sensitivity of the resonant frequency to the coefficient b_1

Figure 3.14

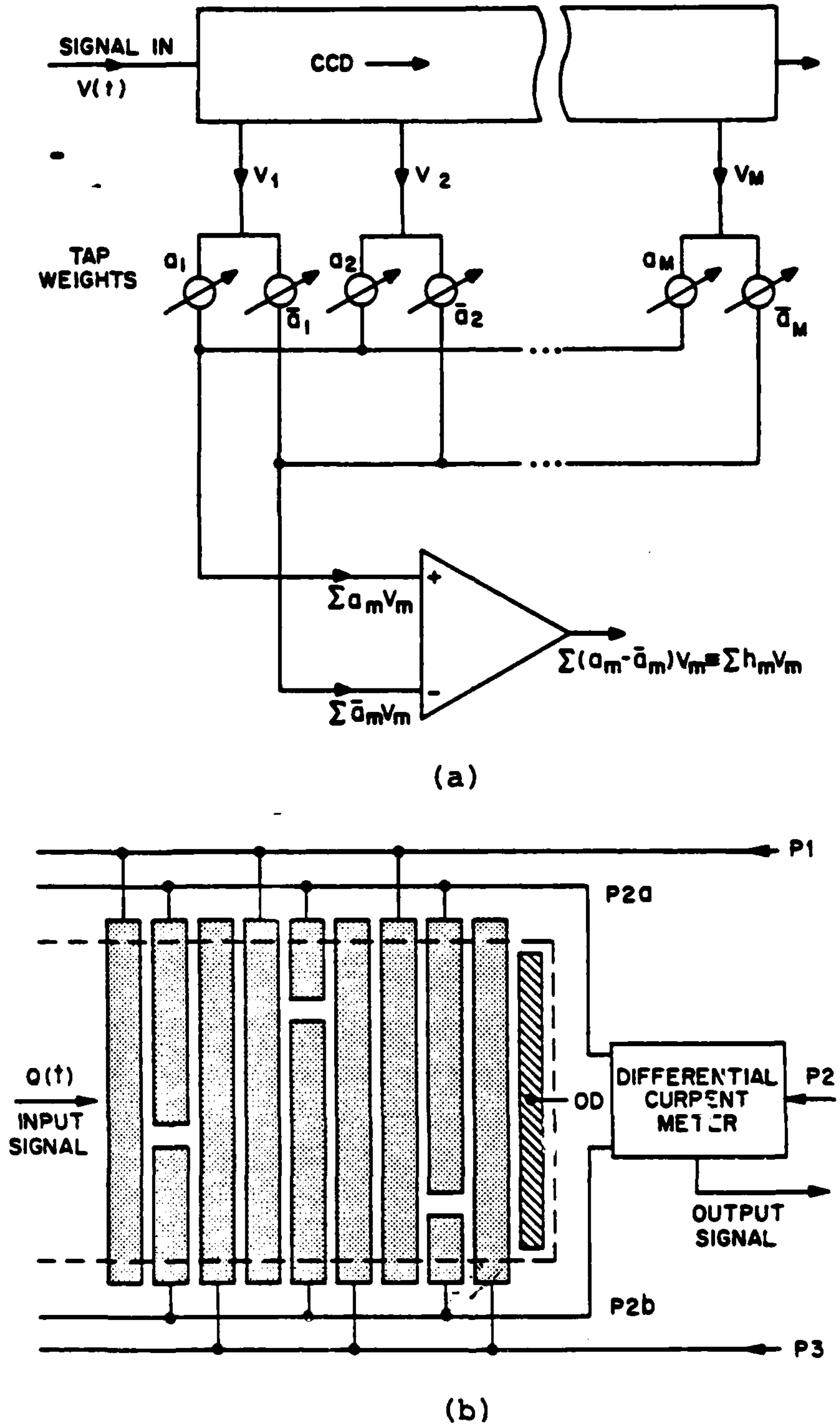
by using split-electrodes which detect the signal charge as it passes under the sense point. The electrode area used for detection defines the tap weight. Figure 3.15 (43) shows the CCD transversal filter approach. These filters are being actively pursued for a variety of telecommunications applications (44, 45), and offer the promise of lower cost, smaller size, and better stability than their active RC counterparts.

Switched-capacitor filters are the newest approach to monolithic filter realization (46-50). In conventional active RC filters, RC integrators such as that shown in Figure 3.16 are required. The time constants for audio frequencies are long and the absolute values of R and C must be precisely set and stable over time and temperature. The switched capacitor approach overcomes these difficulties. Circuit operation, show in Figure 3.16 (47) consists of a sample phase and an integration phase. During the sample phase, the input voltage ($V_1 - V_2$) is used to charge the capacitor C_U. During the integration phase, this capacitor is connected to the integrator, where this difference voltage is scaled and stored on C_I. By switching C_U at a rate f_C, which is high compared to the passband frequencies, the effect is to simulate an equivalent resistance R_{EQ} of value

$$R_{EQ} = \frac{1}{f_C C_U}$$

The integrator gain constant is therefore

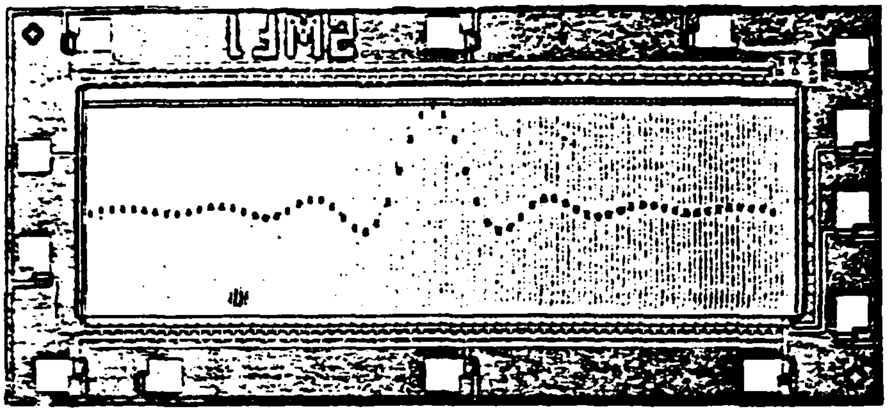
$$w_o = \frac{1}{R_{EQ}C} = f_C \frac{C_U}{C_1} .$$



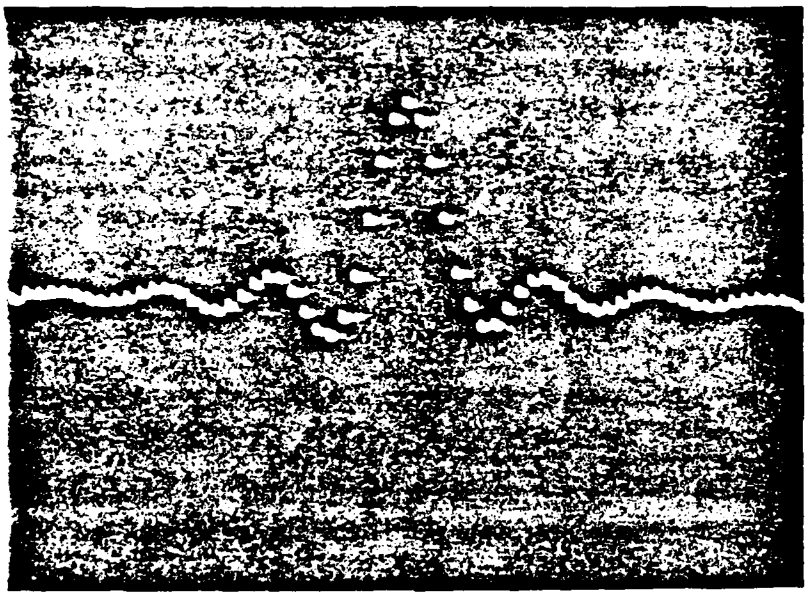
(a) Diagram of a tapped delay line used as a transversal filter
(b) Implementation of a transversal filter using a CCD with split electrodes

Figure 3.15

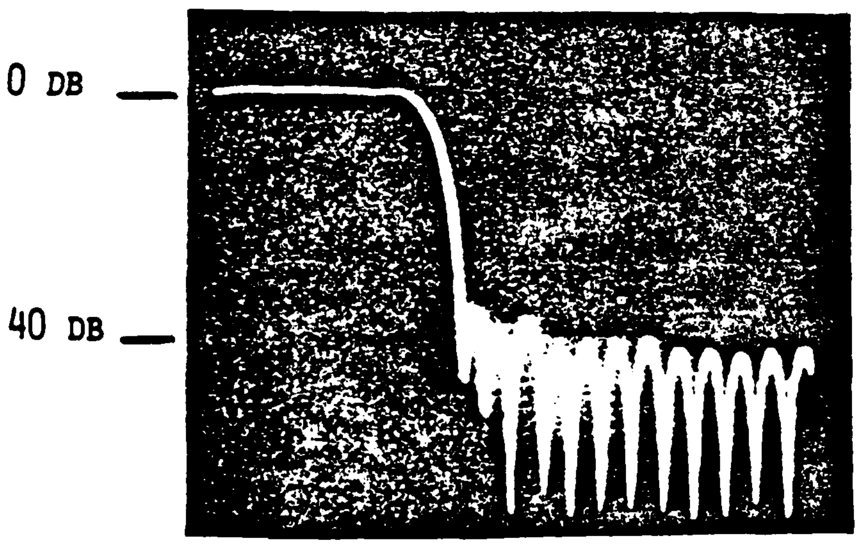
ACUMENICS



(c)



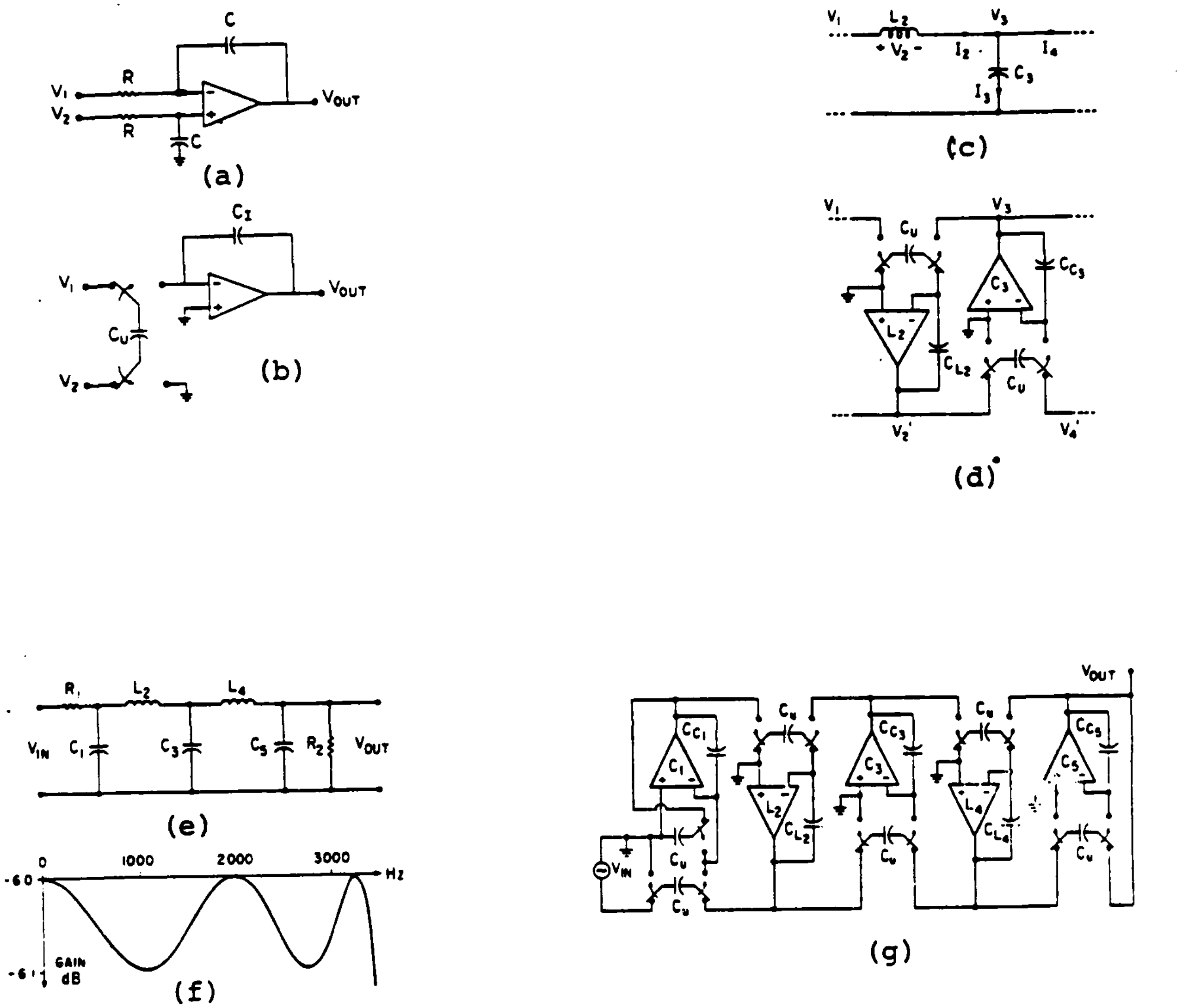
(d)



(e)

(c) Photograph of a low-pass split electrode transversal filter
(d) Observed impulse response
(e) Observed transfer function. The device was operated at 32kHz with a bandwidth of 3.2 kHz.

Figure 3.15



(a) A conventional RC differential integrator
(b) A switched-capacitor differential integrator
(c) A section of an RLC ladder network
(d) Active ladder equivalent from switched capacitor integrators
(e) An RLC doubly terminated fifth-order all-pole low-pass filter and (d) its nominal passband frequency response, and
(g) Switched-capacitor equivalent of the section in (e)

Figure 3.16

The equivalent resistance values which can be realized are very high and very stable, and the circuit performance depends only on the clock frequency and the ratio of two monolithic capacitors. As shown in Figure 3.16, this approach can be extended to the realization of high performance low-cost filters (47). While CCD filters have difficulty meeting the tight requirements on in-band ripple and signal-to-noise ratio found in telephone applications, switched-capacitor filters appear to have fundamental advantages in these areas. It is likely that one or both of these monolithic filtering techniques will see extensive application in telecommunication systems, reducing filter cost by a factor of perhaps five and volume by one-to-two orders of magnitude.

3.7 Line Interface Circuits

A key element in present telecommunication systems is the interface between the input-output device and the switching network. Since the requirements on this interface are primarily determined by the nature of the terminal, these line interfaces will be discussed in this section, even though they presently reside in the switching system itself. In telephone switching systems, their importance is evidenced by the fact that the line/network interface is the dominant cost factor in the system periphery, which in turn is the dominant component of overall switching system cost. For a large space-division electronic switching system in 1970, the periphery outweighed control in cost by a factor of about 1.6:1,

whereas by 1980 for a similarly equipped system this ratio was expected to jump to nearly 3:1, with the periphery accounting for over 50 percent of system cost. It was frequently stated during the early days of solid-state network development that such networks would never be cost effective because even if the crosspoints were free, the line/network interface would be prohibitively expensive compared with those required for metallic (relay) networks. Fortunately, this proved not to be the case.

One possible line interface circuit for use with a space-division solid-state network is shown in Figure 3.17. This interface would interface with the usual "500-set" telephone. Since ringing voltage (about 200 volts) cannot typically be passed through the solid-state network, ringing access is provided by metallic relays between the subscriber line and the ringing bus. Circuitry to terminate ringing when the called party answers is not shown. Power is fed to the telephone via line-feed inductors, which provide a very low resistance at dc for the 48-volt office battery but maintain a high impedance at voice frequencies (about 200 Hz). Call originations are detected by an electronic scan point which senses differential current flow on the line in the off-hook condition. This scan point is implemented as a resistive bridge and comparator. A capacitor blocks the dc line feed voltage from the transformer.

Functional Representation of a Line Interface Circuit
For Use With a Space-Division Solid-State Network

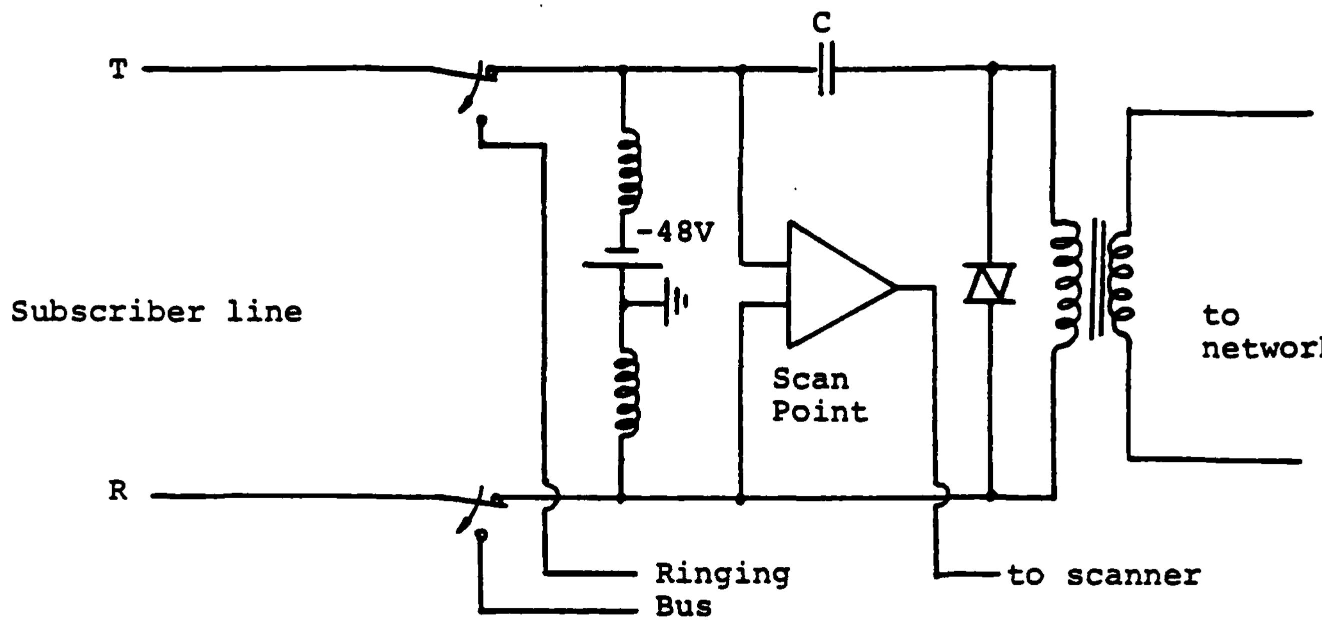


Figure 3.17

ACUMENICS

Signal excursions are limited across the transformer input by a bidirectional clipper diode. The transformer provides a means of coupling bidirectional signal current to the network while blocking common-mode signals on the line. Metallic distribution networks (wire pairs) constitute an effective collection network for lightning bolts and carbon blocks are usually used to limit voltage excursions to a few hundred volts as the lines enter the office. Since telephone lines often run parallel to power distribution lines, inductively coupled common-mode "longitudinal" currents on the telephone lines can be high, sometimes exceeding 100 mA. Thus, the transformer is important in blocking these signals.

While actual line interface circuits may differ in detail, the circuit in Figure 3.17 contains most of the required functions. The line-feed inductors and transformer, and the relays, are bulky and relatively expensive. Nevertheless, they perform their functions well and, combined with the high-voltage/high-longitudinal aspects of the subscriber lines, present a formidible challenge to further cost reduction, miniaturization, or integration. Progress in this area has been relatively slow but is being made.

The evolution of the telecommunication network toward digital terminals, digital time-division switches, and fiberoptic customer lines will have a substantial impact on line interfaces. Adoption of a tone-ringing solid-state telephone would eliminate the problem of ringing access and reduce the line-feed power requirements.

Fiberoptics would eliminate lightning and longitudinals, further simplifying the interface. Thus, these developments have the potential of eliminating most of the present line interface and could reduce the cost of a local switching office by a factor of one-third or more. For existing space-division offices serving conventional telephones, progress is also being made. Floating power converter circuits immune to longitudinals and capable of providing 1000 volt isolation between loop transients and the network have been described (51). Thus, it is likely that some miniaturization and cost reductions will occur even on conventional lines during the 1980s.

REFERENCES

1. R. Grundlach, "LSI in Telecommunications: A Special Report," Electronics, 50, pp. 93-108, April 28, 1977.

2. E. B. Carne, "The Wired Household," IEEE Spectrum, 16, pp. 61-66, October 1979.

3. N. Mokhoff, "Office Automation: A Challenge," Ieee Spectrum, 16, pp. 66-69, October 1979.

4. D. A. Laws and R. J. Levy, "Data Goes Faster, Farther with Chips for Drivers, Receivers," Electronics, 51, pp. 125-129, September 14, 1978.

5. H. Falk, "Chipping in to Digital Telephones," IEEE Spectrum, 14, pp. 42-46, February 1977.

6. D. A. Hodges and V. I. Johannes, eds., "Solid-State Circuits for Telecommunications: A Special Issue," IEEE J. Solid-State Circuits, 14, February 1979.

7. H. J. Hinden, "Technologies, Architectures Compete for Huge Codec Market," Electronics, 51, pp. 105-114, September 14, 1978.

8. Bell Telephone Laboratories, Transmission Systems for Communications, rev. 4th ed., December 1971, pp. 577-582.

9. J. T. Caves, C. H. Chan, S. D. Rosenbaum, L. P. Sellars, and J. B. Terry," A PCM Voice Codec with On-Chip Filters," IEEE J. Solid State Circuits, 14, pp. 65-73, February 1979.

10. C. L. Seacord and D. Vaughn, "Computer Technology Forecast Study for General Aviation," NASA CR-137889, prepared by Honeywell, Inc. for NASA, June 30, 1976.

11. L. N. Heynick, I. Reingold, and A. Sobel, eds., "Special Issue on Display Devices," IEEE Trans. on Electron Devices, 20, November 1973.

12. A. Soble, T. Henion, L. N. Heynick, H. Kressle, B. J. Lechner, and I. Reingold, eds., "Special Issue on Display Devices and and Systems," IEEE Trans. on Electron Devices, 22, September 1975.

13. F. J. Kahn, R. N. Bhargave, W. F. Goede, R. L. Johson, and A. Sobel, eds., "Special Issue on Displays and LED's," IEE Trans. Electron Devices, 24, July, 1977.

14. H. G. Slottow, L. A. Goodman, E. Schlam, and G. B. Stringfellow, eds., "Special Issue on Displays and LED's," IEEE Trans. Electron Devices, 26, August 1979.

15. G. Baur and G. Meier, "Progress in Display-Device Development in Western Europe," IEEE Trans. Electron Devices, 22, pp. 934-940, November, 1975.

16. E. G. Dierschke, "High Efficiency GaAsSb/GaAlAsSb Double-Heterostructure Hemispherical Infrared Emitting Diodes," IEEE Trans. Electron Devices, 26, pp. 1210-1214, August, 1979.

17. R. C. Goodfellow, A. C. Carter, I. Griffith, and R. R. Bradley, "GaInAsP/InP Fast, High Radiance, 1.05-1.3 μm Wavelength LEDs with Efficient Lens Coupling to Small Numerical Aperture Silica Optical Fibers," IEEE Trans. Electron Devices, 26, pp. 1215-1220, August 1979.

18. G. Kramer, "Thin-film Transistor Switching Matrix for Flat-Panel Displays," IEEE Trans. Electron Devices, 22, pp. 733-739, September, 1975.

19. T. P. Brody, F. C. Luo, Z. P. Szepesi, and D. H. Davies, "A 6 x 6 20-lpi Electroluminescent Display Panel," IEEE Trans. Electron Devices, 22, pp. 739-748, September, 1975.

20. I. F. Chang and W. E. Howard, "Performance Characteristics of Electrochromic Displays," IEEE Trans. Electron Devices, 22, pp. 749-758, September, 1975.

21. J. Robert, "TV Image with Liquid Crystal Display," IEEE Trans. Electron Devices, 26, pp. 1128-1133, August, 1979.

22. T. Niina, S. Kuroda, H. Vonei, and H. Takesada, "A High-Brightness GaP Green LED Flat-Panel Device for Character and TV Display," IEEE Trans. Electron Devices, 26, pp. 1182-1186, August, 1979.

23. D. R. Reddy, ed., Speech Recognition: Invited Papers of The IEEE Symposium, New York: Academic Press, 1975.

24. N. R. Dixon and T. B. Martin, Automatic Speech and Speaker Recognition, New York: IEEE Press, 1979.

25. J. L. Flanagan, "Computers that Talk and Listen: Man-Machine Communication by Voice," IEEE Proc., 64, April, 1976.

26. L. R. Rabiner and R. W. Schafer, "Digital Techniques for Computer Voice Response: Implementation and Applications," IEEE Proc. 64, April, 1976.

27. J. Allen, "Synthesis of Speech from Unrestricted Text," IEEE Proc., 64, April, 1976.

28. B. S. Atal, "Automatic Recognition of Speakers from their Voices," IEEE Proc., 64, April, 1976.

29. A. E. Rosenberg, "Automatic Speaker Verification: A Review," IEEE Proc., 64, April, 1976.

30. D. R. Reddy, "Speech Recognition by Machine: A Review," IEEE Proc., 64, April 1976.

31. T. F. Martin, "Practical Applications of Foice Input to Machines," IEEE Proc., 64, April 1976.

32. W. Steffe and D. D. Wen, "Second Generation CCD Line Imaging Devices," ISSCC Digest of Technical Papers, pp. 194-195, February, 1979.

33. N. Koike, I. Takemoto, K. Sato, H. Matsumaru, M. Ashikawa, and M. Kubo, "An NPN Structure 484 x 384 MOS Imager for a Single-Chip Color Camera," ISSCC Digest of Technical Papers, pp. 192-193, February, 1979.

34. D. Powell, "Moving Head Boosts Cartridge Storage," Minicomputer News, pp. 4-6, November 8, 1979.

35. J. A. Rodriquez, "An Analysis of Tape Drive Technology, Proc. IEE, 63, pp. 1153-1160, August 1975.

36. K. E. Haughton, "An Overview of Disk Storage Systems," Proc. IEEE, 63, pp. 1148-1152, August, 1975.

37. P. Snigier, "8-Inch Hard Disk Renaissance Begins," Minicomputer News, pp. 7-13, September 13, 1979.

38. D. B. Davis, "Line Printers -- Market Growth, Composition, and Trends," Minicomputer News, pp. 12-23, October 11, 1979.

39. I. L. Wieselman, "Impact, Non-Impact Printers Offer Mixed Benefits," Minicomputer News, pp. 1-23, October 11, 1979.

40. D. B. Davis, "Serial Printer Markets to Show Growth, Stability," Minicomputer News, pp. 4-13, December 20, 1979.

41. A. H. Bobeck, P. I. Bonyhard, and J. E. Geusic, "Magnetic Bubbles--An Emerging New Memory Technology," Proc. IEEE, 63, pp. 1176-1195, August, 1975.

42. M. F. Tompsett, W. J. Bertram, D. A. Sealer, and C. H. Sequin, "Charge-Coupling Improves its Image," Electronics, 46, pp. 162-169, January 18, 1973.

43. C. H. Sequin and M. F. Tompsett, Charge Transfer Devices, New York: Academic Press, 1975.

44. A. A. Ibrahim, G. J. Hupe, and T. G. Foxall, "Double-Split-Electrode Transversal Filter for Telecommunication Applications," IEEE Journal of Solid-State Circuits, 14, pp. 80-85, February 1979.

45. P. I. Suciu, M. F. Tompsett, J. R. Barner, and P. M. Ryan, "Multiple-Tone Detector with CCD Filters," IEEE J. Solid-State Circuits, 14, pp. 91-96, February 1979.

46. J. T. Caves, M. A. Copeland, C. F. Rahine, and S. D. Rosenbaum, "Sampled Analog Filtering Using Switched Capacitors as Resistor Equivalents," IEEE J. Solid-State Circuits, 12, p. 592-600, December 1977.

47. D. J. Allstot, R. W. Brodersen, and P. R. Gray, "MOS Switched Capacitor Ladder Filters," IEEE J. of Solid-State Circuits, 13, pp. 806-814, December 1979.

49. P. R. Gray, D. Senderowicz, H. Ohara, and B. M. Warren, "A Single-Chip NMOS Dual Channel Filter for PCM Telephony Applications," IEEE J. Solid-State Circuits, 14, pp. 981-991, December 1979.

50. B. J. White, G. M. Jacobs, an G. F. Landsburg, "A Monolithic Dual Tone Multifrequency Receiver," IEEE J. Solid-State Circuits, 14, pp. 991-997, December 1979.

51. L. Freimanis and D. P. Smith, "A Floating Low-Power Subscriber Line Interface," ISSCC Digest of Technical Papers, pp. 180-181, 1980.

Chapter Four

Switching Systems

4. Switching Systems for Telecommunications

The second major element of telecommunications is switching. The switching machines provide the important function of interconnecting the customer terminals and transmission facilities, and increasingly provide the total network with "intelligence", by which we mean the capability to provide a wide range of specialized services simultaneously.

In section 4.1 we will describe briefly the hierarchy of switching systems in the public switched network; in section 4.2 and 4.3 we will describe modern toll and local switching systems; and, section 4.4 will focus on small customer premise switching systems (such as PBX's). Section 4.5 will give a forecast of future developements in switching systems.

4.1 The Public Switched Network

In 1882, almost one hundred years ago, there were 47,700 telephones in service in the U.S. In contrast, in 1978 over 169,000,000 telephones were in service--81% of these were provided by AT&T, 10% by GTE and the remainder by approximately 1500 independent companies and cooperatives. These telephones generated $50 billion in revenues in 1978; included over 250 billon local calls and 25 billion toll calls; provided employment for approximately one million people; and required an investment of $140 billion.

So that all 169 million telephones can talk to each other efficiently and economically, a systematic network has evolved which consists of a 5-level hierarchy of switching centers. Class 5 offices are at the lowest level and are the most numerous. These home in on the next level offices called Class 4 toll centers (of which there are approximately 800 in the Bell System); these in turn home on 230 Class 3 primary centers, 67 Class 2 Sectional centers and ten Class 1 regional centers. Altogether in 1977 there were approximately 20,000 switching offices in the U.S., of which 17,000 were in the Bell System. This formal network structure assures that the longest calling path will not contain more than nine trunks in tandem (1,2).

Voice switching functions are complex. The central office equipment supervises all lines; it detects when a subscriber wishes to make a call (which is commonly known as "off-hook" condition); it collects the dialing (address signaling) digits; it analyzes these digits to determine the type of call and routing; on interoffice calls, it selects the appropriate trunk and passes on the appropriate signaling information; on local calls it selects the talking path; it powers the telephones and detects disconnections; it alerts (rings) the called line; and it attends to billing, administrative and maintenance functions.

In 1977 only two manual central office switchboards were reported in operation; however, 46% of the 20,000 or so telephone company switching centers were of the step-by-step electromechanical types (whose design goes back to 1890), 35% were the more modern crossbar types using relay common control, and 18% were electronic types using stored program control (1). Current industry plans look to the replacement of all electromechanical switches with electronic types within the next two decades.

The advantages of stored program control over electromechanical switching are higher capacity, lower cost,* lower floor space requirements, lower operation and maintenance expense, increased revenues through custom calling services, and greatly increased operational flexibility. (With regard to the latter, the number of features offered for electronic offices has tripled over those offered by electromechanical offices not using stored program control) (3).

---

*AT&T estimates that over the past ten years the cost of a crossbar switch for a typical office has doubled; however, the cost for a similar capacity electronic office has remained approximately the same.

Another feature of the network which is closely related to stored program control is Common Channel Interoffice Signaling (CCIS). In the electromechanicl network, the first prevalent signaling system, used with step-by-step systems, was dial pulse signaling. When common control systems came into use and carrier-type transmission systems became more widespread, multifrequency (MF) tones were used to transmit address information and single-frequency (SF) sets carried supervisory signals over the voicepath circuits used by the subscriber. SF/MF signaling has served well for about 30 years, but it has significant limitations. It is relatively slow, can transmit a very limited amount of information and its capabilities cannot be economically expanded. It is clearly not capable of matching the capabilities of electronic offices and is now being supplanted by the CCIS system (4).

The principle of the CCIS system is to transmit all of the signaling information pertaining to a group of trunks over a separate dedicated channel. All of the signaling interaction between two electronic switches, including network management, can take place between the processors over the CCIS network of voice-grade data lines. The result is a great reduction in call set-up time--one or two seconds on a coast-to-coast call, instead of the previous

ten seconds or more for the former system which utilized the same trunk for both signaling and the talk path. There are many other advantages, special routing information can be carried along with the call. For example, to minimize transmission delay by ensuring there is no more than one satellite hop in the connection, special signals can be transmitted to do this when the call is being set up. Also, messages can be relayed back from any office in the connection to the originating office so that tones and announcements can be applied locally rather than across the network--thus freeing trunks for other calls. The calling party can be identified before answering by a special tone or by an alpha-numeric display. Going beyond the basic CCIS functions for improved network management, AT&T foresees the expansion to tailored stored program control network functions for unique customer requirements, including special inward WATS routing features whereby when a nationwide (800) calling number is used, such calls could be routed to selected regional answering centers rather than to a single center (5,6).

4.2 Toll Switching Machines

Prior to 1970, virtually all toll switching machines were analog metallic crosspoint space-division switches. By analog, we mean that the mode of transmission through the switch was analog, independent of the mode of transmission of the same trunk in the transmission world. By metallic crosspoint, we mean the actual method of establishing the connection was by the use of magnetically

controlled relay contacts. Finally, by space-division, we mean that the means of separating the trunks within the switching machines was by separating them on physically distinct wires, one trunk per wire. The interface between such a switch and the transmission world was very simple, namely a pair of wires carrying an analog signal for each trunk, called a voice frequency, or VF interface.

This type of toll switching machine is perhaps best exemplified by the Bell System's No. 4 Crossbar System, which was the mainstay of the network from 1943 until recently. Between 1965 and 1975 the number of these switches tripled, and in 1976 seven metropolitan areas had four or more of these machines and 22 cities had at least two.

Around 1970, it became obvious that the solid-state technology engulfing the rest of telephony was also the wave of the future in switching networks as well. In particular, engineers started looking seriously at ways of replacing the metallic crosspoints, which were labor-intensive in their manufacture and hence stable or increasing in price, with solid-state crosspoints, which had the potential of following a declining price curve. At about the same time, digital transmission, which was introduced in 1962 in the form of the Bell System T1 system, was expanding to the point of encompassing a significant fraction of the trunks in a toll switching office (particularly the relatively short-haul toll-connecting trunks connecting to local switches).

Solid-state crosspoints could be used in a conventional space-division configuration. However, this had three major disadvantages. First, solid-state crosspoints at that time had a loss variation which was too great to meet transmission requirements. Second, solid-state crosspoints with VF transmission require two pins on the device package for each trunk being switched, and hence the level of integration which could be achieved would be severely limited by pin-out and packaging requirements. Third, this approach did not interface naturally with digital transmission, which was expected to become the dominant mode of transmission in the telephone network.

Using digital transmission internal to a toll switch was a natural way of achieving solid-state crosspoints without the difficulties mentioned above. Obviously the interface to digital transmission would be simple and inexpensive, and there would be no loss variation whatsoever. In addition, if the signals were switched in a time-division multiplexed form similar to that used in the digital transmission world, each integrated circuit pin and each crosspoint could perform the switching function for multiple trunks. Thus, not only could the pin-out problem be solved, but the total number of crosspoints in the switch would be substantially reduced, resulting in an additional economic advantage.

Thus, it was decided in 1970 to develop a digital toll switch, the No. 4 ESS, in the Bell System (7,8). The development took seven years and reportedly cost $500 million, culminating in the first cutover in 1976. The concept of digital switching in which trunks were separated in time (by time-division multiplexing or TDM) as well as space had been investigated in the late 1950's (9), but had to wait for the technology to catch up. The first commercial digital switch (a local switch) was placed in service in France in 1970 (10). The first toll switch was placed in service in England in 1973.(11). In the United States the No. 4 ESS and the Vidar IMA-2 (12) toll digital switches were introduced in 1976, and the Automatic Electric No. 3 EAX was placed in service in 1978 (13). These systems were going into service so rapidly that by December 1978 fully one-eighth of all toll calls in the Bell System were switched by No. 4 ESS (14). A number of manufacturers around the world have toll digital switching machines under development, as summarized in [15]. It appears that practically all metallic crosspoint toll switching machine developments have been abandoned in favor of digital switching approaches.

4.2.1 Architecture of Toll Digital Switching Machines

A simplied block diagram of a typical toll (Class 1 through 4) digital switching machine is shown in Figure 1, and a detail of the network is shown in Figure 2. The standard interface to the switching network is the DS1, which is the slowest speed bit stream in the digital transmission hierarchy.

It consists of 24 VF trunks time-division multiplexed together, with a total bit rate equal to 1.544 Mb/s. DS1's originating in the transmission world can therefore be interfaced to the switch directly as shown in Figure 1. There is also considerable analog transmission in the intertoll network which is in frequency-division multiplexed form. The lowest signal in this hierarchy is the group, in which 12 trunks are multiplexed together. A transmultiplex, which converts two FDM groups into a single TDM DS1, efficiently interfaces analog transmission facilities to the digital switch (16). Finally, every office has a small number of service circuits and special services (such as private lines) which can be interfaced through a conventional channel bank such as the Bell System D4.

The cross-office paths are set up through the switching network under control of a stored-program control (SPC) processor, which is essentially a special-purpose computer. The processor communicates

TOLL DIGITAL SWITCH AND TRANSMISSION INTERFACES

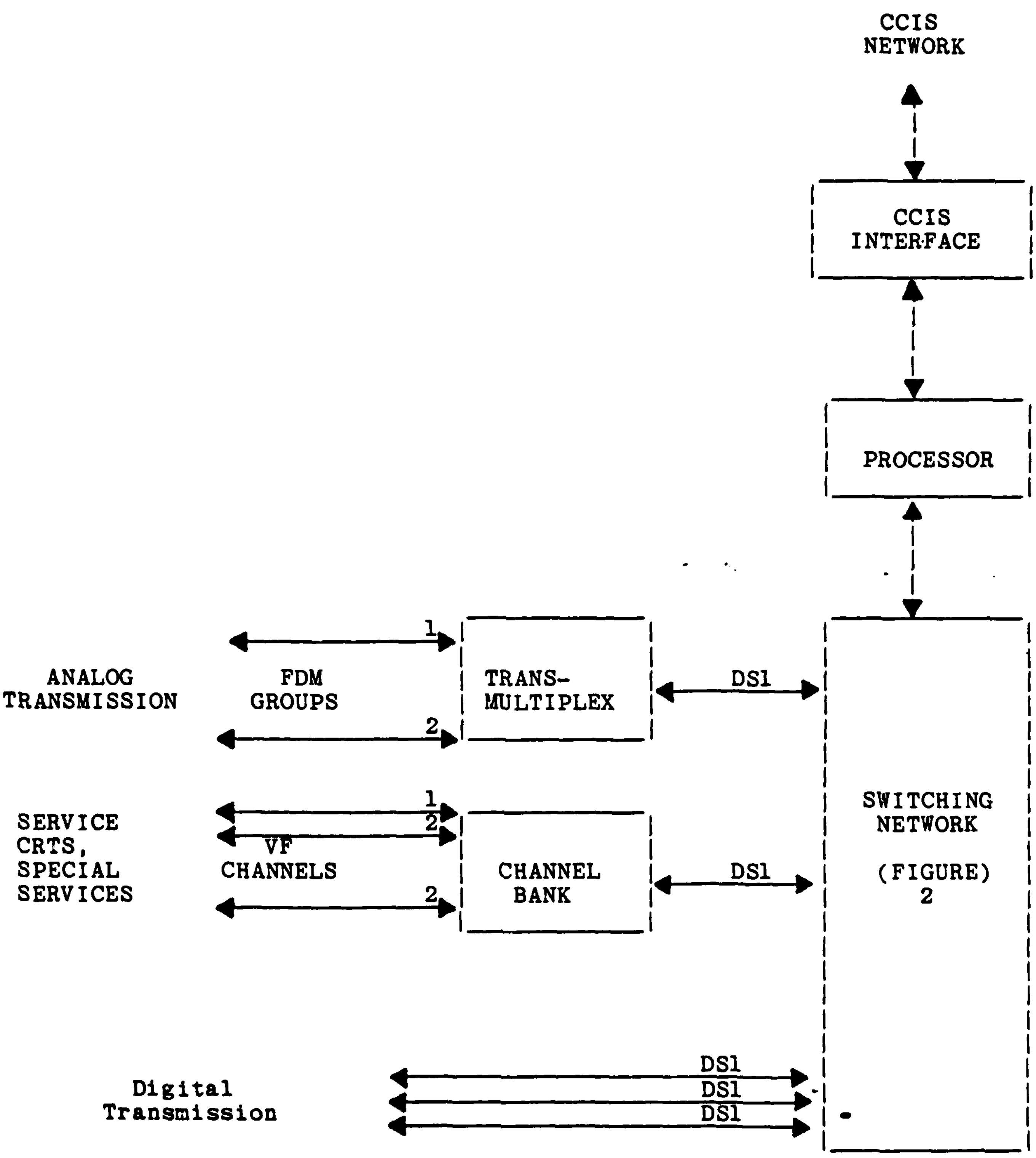


Figure 4.1

TYPICAL TOLL DIGITAL SWITCH NETWORK

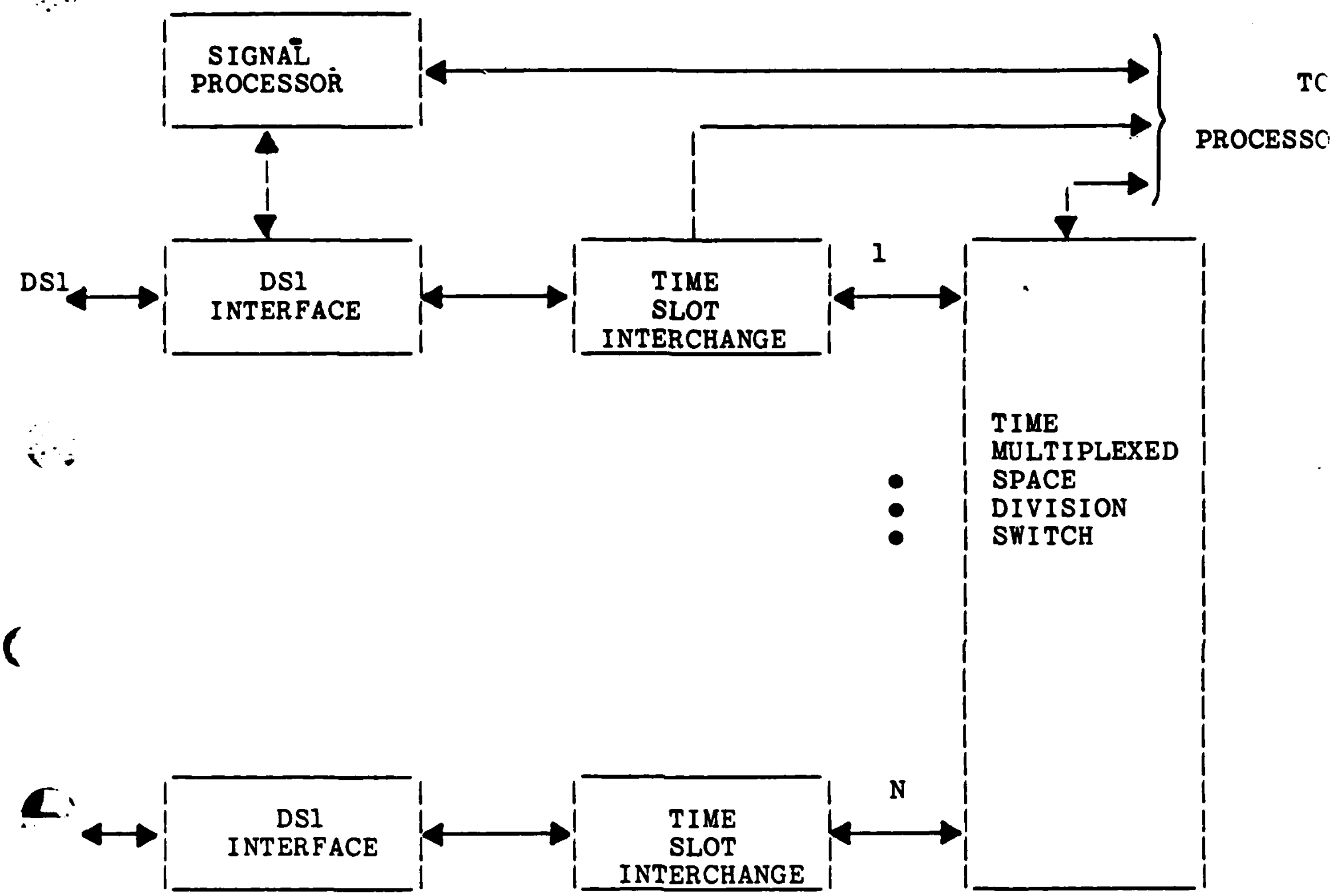


Figure 4.2

with other switches to set up a lengthy telephone connection in one of two ways. One is on the trunks themselves, and this signaling path is interfaced to the processor by way of the network. The other signaling method, which is expected to become ubiquitous, is common-channel interoffice signaling (CCIS). It is esentially a special data network overlayed on the telephone network for signaling purposes. The processor must therefore have a separate interface to the CCIS network.

The heart of the switch is the network, which is shown in simplified form in Figure 2. As shown in Figure 1, the standard interface to the network is the DS1. The DS1 interface block performs many functions relating to transmission, signaling, and switching. First, it performs such transmission functions as bipolar-to-unipolar conversion, framing, and violation monitoring. Second, it strips off the signaling bits, indicating the on-off hook status of each trunk (which is included in the bit stream by robbing information bits in each sixth frame). Finally, it has a buffer store which absorbs any difference in the phase of the internal switch clock and the timing reference of the incoming bit stream. At the output of the DS1 interface block, the 24-channel frames of the bit streams are all lined up together, this being essential for the subsequent switching function.

The signaling bits which are stripped off the incoming DS1's in the DS1 interface are all routed to the processor indirectly through another unit called the signal processor.

The switching function itself is performed by the time-slot interchange (TSI) and the time-multiplexed space-division switch. Together, these units must be able to route a sample from one trunk a) to any other DS1 and b) to any time slot on that DS1. Thus, the TSI enables the sample to be shifted from one time slot to another, by simply reading all samples into a buffer store in the order that they appear on the incoming DS1, and reading them out in the next frame in a different order. The space division switch then switches the sample in its new time slot to any of the other DS1's (or possibly the same one). Since the space-division switching function is different in each time slot, the space-division switch reconfigures itself in each time slot. One major economy is the fact that the space-division switch is performing multiple duty, completely reconfiguring in each time slot, thereby substantially reducing the ratio of crosspoints to trunks.

There are many variations on the architecture of Figure 2. For example, it is possible to reverse the order of time slot interchange and space-division switching. It is also common to time-division multiplex up to a higher bit rate prior to the time

slot interchange in order to more efficiently utilize the space-division switch. As an example, in the No. 4 ESS five DS1's are multiplexed together (to 120 trunks) prior to time slot interchange, and the space division switch is 1024 by 1024. The total number of trunk terminations is therefore 120 times 1024, or 122,880.

4.2.2 Advantages of Digital Switching

There are numerous reasons why digital switching has become ubiquitous in new switching developments, both toll and local. Perhaps foremost at this time is the low cost of the switching network itself, due to the solid state crosspoints and the fact that a relatively small network is time-division multiplexed over a much larger number of trunks. In addition, every indication is that the digital logic technology (such as ROMs, RAMs, and multiplexers) extensively used in the digital switching network is likely to rapidly decrease in cost, whereas competitive metallic crosspoint approaches are labor-intensive and hence susceptible to inflationary cost increases.

Aside from the switching network itself, the interfaces to the transmission world in a toll digital switch are much more natural and cost-effective than in a space-division switch. This is because the incoming digital tranmission trunks are in the same

format as is used internally in the switch network, and digital transmission constitutes an ever-expanding portion of the trunk terminations. In contrast, the VF metallic trunk interface to a space-division switch is almost non-existent in the transmission world. This desirable coordination of transmission and switching internal formats is known as integrated digital transmission and switching (17).

Another factor which is not insignificant in the advance of digital transmission and switching is the simple way in which signaling can be accomplished in the digital transmission format (by simply multiplexing extra bits into the bit stream for this purpose). In contrast, for any other combination of transmission and switching, expensive in-band modems are required for signaling (such as Signal Frequency or SF signaling).

The time-division multiplexed digital interface to a digital toll switch also results in some economies in the realization of certain signal processing functions. A single functional unit placed at this interface can be time-shared over multiple trunks, with a hardware savings as compared to replication on a per-trunk basis, and there are also maintenance advantages to digital implementations. An example is the echo-supressor terminal for No. 4 ESS (18) which has a ten-to-one cost advantage and a 25-to-one space advantage over the analog device it replaces. Digital switching allows these savings regardless of whether the transmission trunks are analog or digital.

There are administrative and maintenance advantages to digital switching. Digital hardware is more amenable to automated fault recognition and isolation procedures, resulting in savings in recurring costs such as salaries. In addition, because the digital switching network is so inexpensive, it can be made virtually non-blocking with small economic penalty, resulting in the elimination of many expensive traffic measuring and load-balancing administrative procedures.

All the advantages we have cited make toll digital switching economic in today's telephone network, even in the presence of significant analog transmission. Perhaps the greatest advantage of digital switching is the flexibility that it provides for the future. Once various types of signals (such as audio, video, data, etc.) are converted into digital form, they can be readily accommodated in a digital transmission and switching network. Since non-voice services are expected to become ever more prevalent in the near future, the advantages of entering this era with a flexibile integrated digital transmission and switching network are overwhelming.

Thus far we have concentrated on the considerations in adopting digital switching. The conversion to stored-program control (SPC) of switching machines (independent of whether the network is analog or digital) occured much earlier and also has important consequences for the future. The SPC switches are much more flexible in the

features they can provide, and allow those features to be changed and added over the lifetime of the switch. The conversion to CCIS signaling in conjunction with SPC also results in a dramatic increase in the capability to provide new services, since the communication capability between switching machine SPC processors is greatly enhanced.

Of course digital switching also has disadvantages. One is the cost of conversion to a radically new technology. Perhaps the greatest new technical and administrative burden introduced by digital switching is the need to synchronize the switching machines together (as was mentioned earlier, it is assumed that the incoming DS1's are synchronized to the switch internal clock). Rather elaborate network synchronization mechanisms have been set up to accommodate this need for synchronization (19).

4.3 Local Digital Switching Machines

Local (Class 5) switching machines have undergone an evolution similar to toll switching machines. Until very recently, the predominant architecture was a space-division metallic cross-point switch with stored-program control. In contrast to toll switching, the first break from this architecture was by a group of relatively small telephone equipment manufacturers who foresaw

the application of digital local switching, particularly to smaller switches in rural areas. The large manufacturers have since followed suit with local digital switching developments of their own, and it appears that as in the toll arena digital switching machines will become ubiquitous.

The earliest application of local digital switching was in France in 1970 (20), which now has more than 500 thousand lines in service. Starting with the Stromberg-Carlson DCO in 1977, several small local digital switches have become available in the U.S., and two large local digital switches are under development at Western Electric and Automatic Electric (15).

A common architecture of these machines was anticipated in 1959 (9) and is shown in Figures 3 and 4. The switching network of Figure 3 uses time-division and space division techniques similar to toll switches. The interfaces to transmission trunks are similar to toll switches also, with the single exception that there are essentially no local or toll-connecting FDM trunks and hence transmultiplexers are not required. Likewise, the processor and CCIS interface are similar to those of toll switches.

The major distinction between local and toll digital switches, as well as between digital and metallic local switches, is on the subscriber line side of the switch. The line interface switch

LOCAL DIGITAL SW...

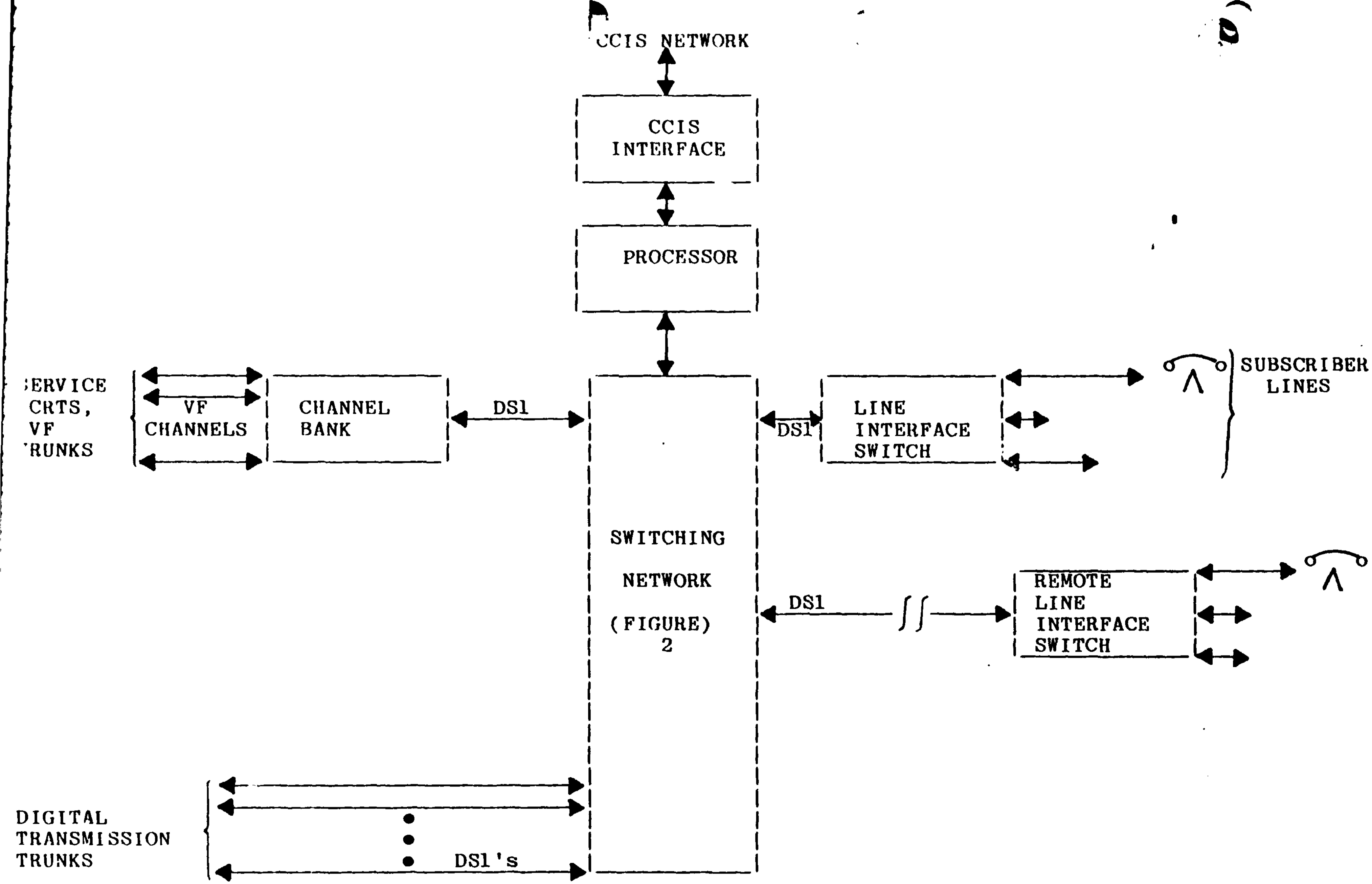
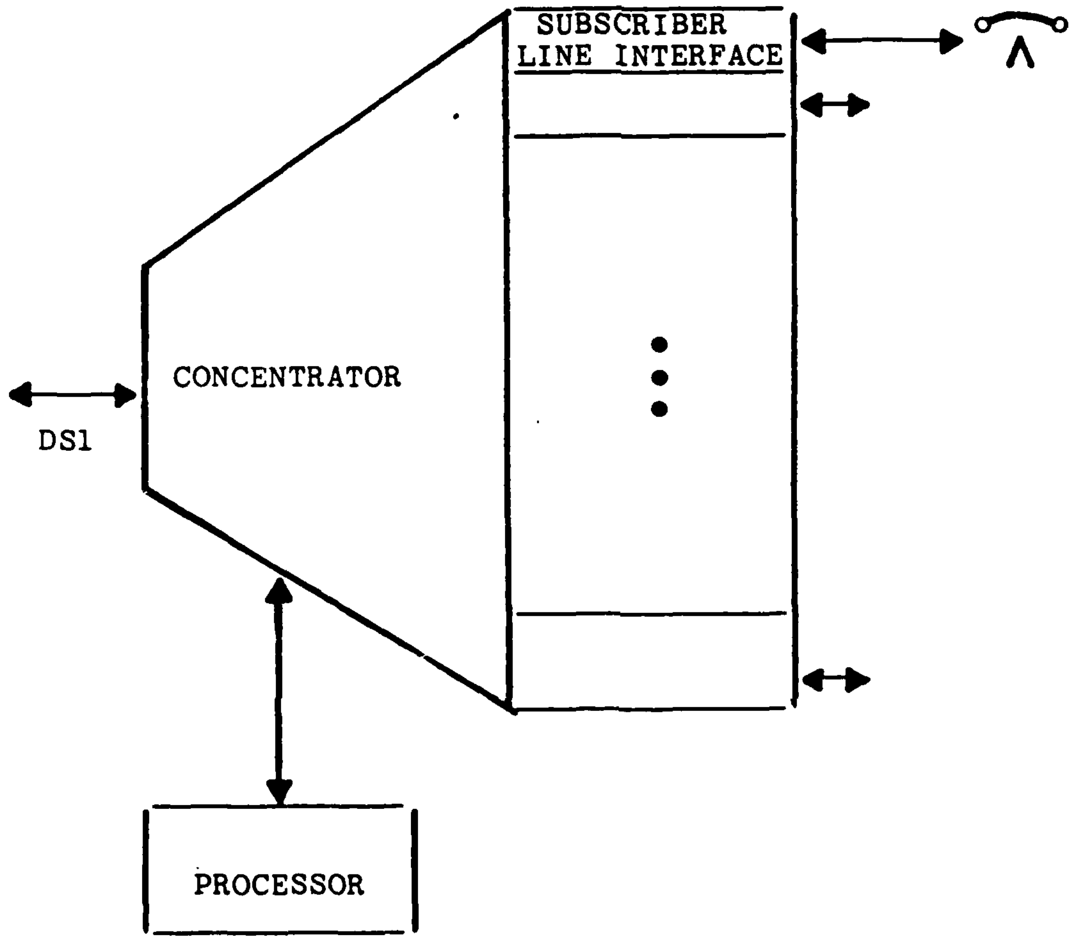
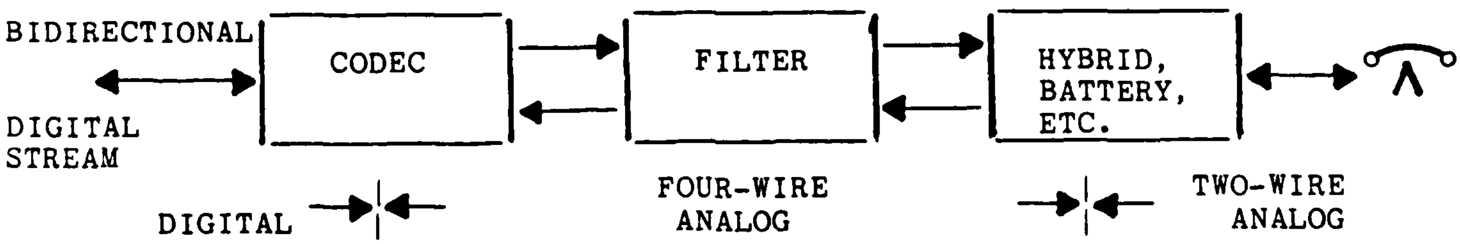


Figure 4.3

LINE INTERFACE SWITCH



(a) ARCHITECTURE



(b) DETAIL OF SUBSCRIBER LINE INTERFACE

Figure 4.4

interfaces the subscriber lines, which are metallic, to the DS1 interface to the digital switching network. An important feature is this standard DS1 interface, since it allows the line interface switch to be located remotely from the switching network and connect to it by standard digital transmission facilities.

The architecture of a typical line interface swith (local or remote) is shown in Figure 4. An important function is to interface the metallic two-wire subscriber loop to the digital transmission format. This subscriber line interface circuit is shown in block diagram form in Figure 4b, where we see the important functions required. The two-wire loop must be powered (battery) and converted to four-wire transmission (hybrid). Then in the direction of the switch, the signal is low-pass filtered to eliminate aliasing distortion in the subsequent sampling, and then converted to digital in the coder. In the direction of the subscriber, the signal is converted to analog in the decoder and then reconstructed with another lowpass filter.

The subscriber line interface just described is an important part of the line interface switch of Figure 4a. The other primary function is concentration, usually by a factor of six or seven, in recognition of the low connect time of the average subscriber loop. The concentrator is controlled by a local processor, whch also provides maintenance features.

The line interface switch processor communicates with the base switch processor over the DS1 interface. The two processors can then cooperate in the setting up of a connection. Some line interface switches also provide the capability to switch a call between two subscribers on the same line interface switch entirely within that switch, thereby freeing up capacity on the link to the base switch and also providing the capability for a modicum of service in case of failure of the transmission facility to the base switch.

4.3.1 Considerations in Local Digital Switching

The economics of local digital switching are not quite so favorable as in toll switching, particularly for the larger switches. This is because the interface to the subscriber line is not as inexpensive as the interface to the digital transmission trunk. In a local digital switch, it is therefore the cost of the subscriber line interface which predominates.

In an analog space-division local switch, the only function which has to be performed on a per-subscriber line basis is supervision (determining if the phone is off-hook). Other functions, such as loop testing and ringing can be performed

on a service circuit basis through the metallic crosspoint network. In contrast, a digital switch cannot handle the large voltages required for these functions in the digital network, and consequently a metallic high voltage access must be provided external to the digital network. Similarly, none of the functions of Figure 4b are required in an analog metallic crosspoint switch on a per-line basis.

Although the subscriber line interface is more expensive for a local digital switch, there are compensating savings. The interface to digital transmission trunks is less expensive. Also, the remote line interface switch is very useful in reducing the number of copper wires in the subscriber loop plant. In rural areas, small central offices can be replaced by remote line interface switches connected to a larger base switch, thereby centralizing administration and maintenance and reducing trunking requirements.

The economics are more favorable for small local digital switching because the digital switching network is a dominant part of the cost and is less expensive than for an analog switch. In larger digital switches, the per-line cost of the subscriber line interface dominates. One modification of the architecture of Figure 4a which looks attractive, particularly for the larger switch, is to place the concentration between the subscriber line and the subscriber line interface. The concentration must then be analog, either

solid-state or metallic. The number of line interfaces is then dramatically reduced, with the resulting savings partially offset by the increased cost of the analog concentrator.

Another modification of this basic architecture which shows promise of both cost savings and the potential for new services is the digital telephone instrument. In this configuration, the codec would be placed in the telephone instrument itself, and the subscriber loop would employ 64 kb/s or higher speed digital transmission. Supervision and signaling could be performed over the digital subscriber loop more naturally than by using high voltages and tone signaling as at present. The present subscriber loops can support this digital transmission out to 5-8 km without repeaters. Higher bandwidths and/or longer distances may require fiber optics technology. The first digital telephone set is presently being test marketed in Denmark (22).

4.4 Small Switching Systems

The evolution in small switching systems for use on customer premises (such as Private Branch Exchanges or PBXs) is occuring perhaps more rapidly than anywhere else in the telephone network. These systems involve relatively little development effort, and are therefore accessible to a number of small and innovative manufacturers.

4.4.1 Small Telephone Systems - A Historical Perspective

The movies of the 1920's frequently showed a harried editor of a newspaper frantically attempting to cope with a half dozen phones on his desk, all ringing at the same time. At that time, there were no organized, small, multi-line telephone systems for use on the customer's premises. The first simple attempts to provide some control was the installation of a switch on a phone that looked like a key so that one could switch from one line to another--hence, the term "key telephone system" (KTS) has been carried on for five decades as being synonymous for small telephone systems that could receive and hold several calls simultaneously.

To meet customer requirements, telephone companies in the 1920's supplied what we now call key telephone systems by designing stations, relays and wiring plans for each customer. Some of these plans were standardized, but they lacked flexibility, making installations and changes difficult and expensive. In 1938 the Bell System developed the first standardized key telephone system, the type 1A, using building blocks called key telephone units (KTUs) or line cards, which consist of relay circuits for holding calls and supplying visual and audible signals. The system functions are controlled by pushing buttons that are equipped with lights to indicate system status. An array of buttons is called a keystrip and a telephone so equipped is called a key telephone (23).

The basic purpose of a key telephone system is to permit the user to pick up any of several lines and to hold calls associated with any of these lines. A typical key telephone system consists of several telephones which have access to a subset of four to six lines. System configurations can range from one to perhaps one hundred telephones (which are also called stations). Various types of station sets are used with key telephone systems.

The most common can handle up to 6 lines, and there are various types of consoles that handle up to 30 lines. The common line states and corresponding lamp indcations are shown below:

Key Telephone Lamp Signals

Line State	Lamp Indication
Idle	Not illuminated
Busy	Steady illumination
Ringing	Flashing - 60 flashes per minute
Hold	Winking - 120 flashes per minute

It is estimated that about one half of all installed key telephone systems work into PBX's, which in turn connect to central offices, and the other half are connected directly to central offices.

In addition to the ability to pick up and hold lines, a key system can provide a variety of other features, such as conferencing, intercom service and toll restriction. A conferencing arrangement allows a station user to connect three or more parties simultaneously. Intercom service provides additional channels on which stations can intercommunicate without the need to place a call through the PBX or the serving central office. In manual intercoms, all stations have access to a common talking path, and signaling is done with a buzzer. Dial intercoms generally have more than one talking path, and signaling is done by dialing. Toll restriction circuitry prevents selected stations from making toll calls.

The type 1A1 KTS was introduced in the Bell System in 1953 and considerably reduced the installation effort in relation to that required for its predecessor. The type 1A2 utilities modular packaging, making use of solid-state circuits and plug-in KTUs. Electromechanical relays used in the earlier KTUs are now being replaced by triacs (PNPN diodes) with attendant space and power savings (24). The system design concept for the type 1A2 KTS provides for considerable flexibility and feature offerings and it is also offered by a number of different manufacturers under Western Electric licenses, including ITT, Northern Telecom, GTE-Automatic Electric, Stromberg-Carlson and San-Bar.

Another step in the evolution of key systems is repesented by the Bell Com-Key® systems (which can handle from 4 to 21 lines and from 16 to over 50 stations). These systems were introduced in 1975 as the result of the competition in the telecommunications equipment market initiated by the FCC's Carterfone Decision in 1968, resulting in lower-priced, factory wired and tested systems that could be installed and maintained with less effort than previous systems. In the case of the Com-Key® 718 system, each telephone uses a 25-pair cable, has ten buttons for incoming lines, intercom lines and hold, and is available with rotary or tone dialing. To simplify installation, each incoming line appears on the same button position on all phones, making life easier for the installer. Some of the features include: voice signaling (allows the attendant to speak to someone via the system intercom without the called party having to pick up the handset), tone signaling (supplies two different tones to identify inside or outside calls), music-on-hold, paging, privacy (prevents other persons from listening in on a conversation), privacy release, and power-failure transfer (maintains basic but limited service during a commerical power failure) (25).

Microprocessors are being introduced to small telephone systems. In Bell's recently announced Horizon® system (26), the microprocessor makes use of a read-only memory (ROM) that contains the system program--the routines that dictate system features and operation.

Information is entered into this memory during the manufacturing process and cannot be changed. The other memory, a random-access memory (RAM), contains information about the status of the system and holds translation information such as line and station numbers, button assignments and station restrictions--information that can be changed through an external access unit. (Previously, it was necessary to make hard-wired changes to enter translation data into the older systems).

Instead of using the 25-pair cables between the telephones and common equipment of the older key systems, the new electronic systems now on the market make use of one, two or three pairs of wires running from the telephones to the common equipment with consequent great savings on premise writing requirements. This reduction in premise wiring is made possible by the transition from relay logic used in the older systems to the use of micro processor logic.

The new key systems that make use of microprocessors have built-in maintenance diagnostics, which can automatically test the memories and control circuitry. In the case of failure, light-emitting diodes (LEDs) glow at the central answering position and in the control unit on the malfunctioring circuit packs. (Installed prices for key telephone systems when purchased ..y from about $500 per station for 1A2 systems (which use electromechanical

relays) to perhaps as much as $1000 per station for types making use of microprocessor technology. (The Bell System, since its inception, has only leased terminal equipment to its subscribers. At the present time, some telephone companies provide their subscribers with lease or purchase options. Since 1969, as the result of the Carterfone Decision, customer ownership of telephone terminal equipment has been permitted throughout the U.S.).

4.4.2 Private Branch Exchanges (PBXs)

The basic function of a (PBX) is to connect PBX stations to each other or to trunks to the serving central office. In early manual PBXs, the attendant performed all the inter-connections, incoming as well as station-to-station, at a cord-type switchboard.* Most modern PBXs are dial systems and the interconnections are accomplished automatically.

The trunks that connect the PBX to the central office are fewer in number than the PBX stations as it is rare that all PBX stations will attempt to talk outside of the PBX at the same time. Aside from the normal trunks to the central office, a wide variety of private line circuits can be connected to a PBX, such

*The No. 4 Manual cord-type board (1904) was the first standard PBX in the Bell System (27).

as off-premise stations (OPS), tie trunks (which connect to other PBXs) and foreign exchange (FX) trunks (which provide direction connection to a centrol office other than the normal local exchange).

PBX installations can vary from just a few stations to several thousand. In reality, there is no clear-cut distinction between small PBXs and key telephone systems. As indicated earlier, large PBX configurations frequently will include the connection of several key systems.

The PBX-type service can also be provided by connecting all stations directly to the serving central office in what is known as the CO Centrex system. In such a system, there is no switching equipment on the customer's premises and it is frequently cost effective when the customer is located a short distance from the central office. (There is a looplength dependent cost penalty as the distance from the central office increases). Centrex features offered are similar to those offered by modern PBXs.

In a PBX there is a matrix of switches that effect the connection desired and a means of control, which can be instructed by the user to cause the switches to operate. PBX systems over the years have evolved around three basic switching concepts involving space divison, time division and frequency division multiplexing (28).

Examples of space division switching include cord-type manual switchboards and automatic electromechanical switches, including the Strowger step-by-step switch and other types of two-motion switches, relay and reed switches and the crossbar switch. The earlier electromechanical switches made use of wired logic for control purposes.

In the 1960's and early 1970's several PBX designs evolved using reed and crossbar switching networks which made use of electronic common control (i.e., stored program control) methodology. With stored program controlled processors, translation changes, new service additions, and other modifications can be implemented as changes in the program, rather than making complex hardware and wiring changes.

Typically, these processors interpret and execute, one-by-one, the instructions of the program; they contain a memory that stores the program; they have a memory that functions as an erasable scratch pad to record and accumulate data during call processing; they have scanners through which the central control receives information such as station on-hook, off-hook, dialed digits, etc.; and they have built-in signal distributors through which they cause network switches to operate. Most stored program processors operate in a time sharing mode whereby many calls in various stages of completion are handled simultaneously. Systems using stored program control, therefore, exhibit much more flexibility and versatility than those using wired logic.

The transmission medium in a time-division switching matrix can be either analog or digital. Those PBXs using sampled analog technology make use of either pulse amplitude modulation (PAM) or pulse width modulation (PWM); while those using digital formats use pulse code modulation (PCM) or delta modulation.* As with space division systems, control of time division systems can be achieved through the use of wired logic, stored program control microprocessors, distributed microprocessor techniques, or in the [illegible] of some large systems, minicomputers can be used.

---

*One manufacturer, Digital Telephone Systems, is currently [illegible]ems making use of delta modulation.

In frequency division switching, a particular RF carrier is transmitted over a coaxial cable to the receiving telephone which translates the modulated carrier to voice signals. Switching, therefore, is accomplished by tuning to the appropriate channel.

Although experiments with electronic switching began in the late 1940's, it was not until the 1960's (with the development of the semiconductor and computer industries) that functional telephone switching systems were developed--such as the Bell 101 ESS (a time division switch) and the Bell 800-series PBXs that used transistor-resistor logic control of crossbar switches. It turned out that the switching matrix proved to be the most difficult section of the system to reduce to electronics as the earlier semiconductor devices were not as good as metallic contacts for switching purposes. As a consequence, a considerable number of PBXs on the market today feature electronic common control of electromechanical switches. In fact, of the 200 or so PBXs available in 1979, three fourths use space division (mainly electromechanical)* switching and a little more than half make use of electronic common control circuitry.

---

*A rare example of a system using semiconductor crosspoints is the ITT TE-400 series PBXs which use a PNPN diode switching matrix.

Several major studies are required in the selection of a PBX: traffic expectations, required features, economic considerations, and reliability requirements.

Traffic engineering is a highly specialized field that attempts to assure that calls will not be delayed excessively when being processed by the switch. "Grade-of-Service" is a probability statement related to a specified traffic handling capacity to indicate the expected proportion of calls that would be blocked or not completed.

A typical industry measure for traffic handling capacity is <u>hundred-call-seconds</u> (CCS). Examples: one CCS is equal to one call lasting one hundred seconds; or a 1-hr call (3600 seconds) equals 36 CCS. The traffic capacity of a system is measured in CCS/line and therefore equals the calling rate multiplied by the holding time in seconds divided by 100; where the calling rate is defined as the number of originating and terminating trunk calls plus intercom calls during the busy hour of the day, and the holding time is the average duration of calls in seconds. Benchmarks for CCS/line ratings for PBXs are:

Heavy traffic	-	8.0
Medium traffic	-	5.0
Light traffic	-	3.5

As an example, for a Grade-of-Service figure of P=0.02 (or P.02), at a given CCS/line traffic handling capacity, one would expect 2% of all calls to be blocked. (Most systems are rated at P.01 or P.02). For example, Bell's 101 ESS in its maximum configuration (576 lines and 332 trunks) is rated at 9.0 CCS/line with a Grade-of-Service of P.02. And Bell's Dimension® 100 PBX, at its maximum configuration (100 lines and 22 trunks) is rated at 16 CCS/line with a Grade-of-Service of P.01 (29).

With regard to required features, the needs of the user must be taken into consideration. Users fall into two classes: large users with competent technical and administrative communications staffs and smaller users without such resources. A further break down of users illustrates how basic system packages for each group tends to be unique:

- Industrial, financial and government groups are heavy users of data communications, private networks and both large and small communications sytems.
- Service providers such as hospitals, educational institutions and the lodging industry require sophisticated administrative communications and communications that encourage the use of their services.
- Retail and transportation users require fast and efficient communications typified by automatic call distribution (ACD) systems (30).

It becomes apparent that there is no such thing as a universal PBX, covering the range from 10 to 10,000 lines.

Over 400 features, many of them software options, are now available; however, only 40 or 50 operating features appear to have universal need among all end user groups.* Some of the interesting new feature options include "electronic" telephones, traffic analysis and control, automatic call distributors, software options, voice digitization applications, electronic mail, and electronic tandem switching (networking).

Even telephones are undergoing substantial changes. Stations used with Bell's Horizon® key system and Dimension® PBX make use of tone ringers instead of electromechanical bell and solid-state red and green LEDs as visual indicators instead of the previously used filament lights. Other manufacturers, notably ROLM and Northern Telecom, have gone further and are providing sets with alpha-numeric displays, additional control functions and loudspeakers. Some of the new phones have such functions as call timers, clocks, identification

---

*Furthermore, the wide variety of system options on advanced systems often overwhelm many users with the result that full system capability is rarely attained. Special training programs are often required to indoctrinate personnel in the use of system features.

of the calling number* and equipped with call forwarding, speed calling, call back, do not disturb and conferencing modes. The new phones with such features are expensive, some having price tags reportedly approaching $800.00 each.

Traffic analysis and control can take on almost anything a customer may require for traffic engineering studies, accounting and control purposes, including sophisticated "station message detail recording," automatic (usually least-cost) route selection and toll restriction services.

Automatic call distributor (ACD) functions which include "delay" recordings -- such as, "Wait for the next available operator --" and a means of load distribution to each operator position were originally furnished as stand-alone units. As ACD terminals have demonstrated a position in the market place, PBX manufacturers are now enhancing new systems with ACD options. As ACD features are relatively mature, and are not expected to change much, the areas of system control and monitoring of the ACD function, however, are expected to undergo further refinement.

---

*With the eventual extension of Common Channel Interoffice Signaling (CCIS) features to all telephone company electronic switches, this feature could be made available nationwide as an option to all users of the public switched network--and this could take place within the next two decades.

Software packages for electronic PBXs generally provide for call processing, maintenance diagnostics and administrative functions. The latter are designed to reduce the effort required to add or rearrange features and services provided by the system--including rearranging service assignments for telephones without having to rewire the premises as was required in the pre-electronic processor era. Special software packages are also available. For example, Bell released the Electronic Tandem Switching (ETS) package in October 1978 for use with the Bell Dimension® 400 PBX. This package was particularly designed for customers with substantial interpremise communications. The major features include a uniform numbering plan for all phones in the customer's private network; a queueing feature which functions primarily in the peak busy hours; automatic (least cost) route selection; facility restriction levels (authorization codes are required for use of certain facilities); automatic circuit assurance (self-monitoring of intercity services); station message detail recording; and customer administration of the system, including: assignment of employee calling privileges, override of time-of-day routing patterns, activation and deactivation of trunk group queues, facility assurance reports, station rearrangement and changes in individual station lines, changes in hunting and pickup groups, changes in signaling options, changes in station numbers, and automatic polling of traffic data and network status data.

Cost seems to be the dominant market factor for the smaller PBXs, and this may be the reason that only 15% of the systems of 100 lines or less currently make use of time division multiplexing and only 30% of this class use electronic common control. On the other hand, half of the systems currently on the market between 101 and 200 lines use time division switching and two-thirds of this group use electronic common control (7), apparently indicating that larger and more affluent users are willing to invest in PBXs having advanced features. Installed costs for PBXs can run from about $500 per station for the smaller and older technology systems upwards to $1500-$2000 per station for the new systems with advanced features.

System reliability and maintainability is of great importance to many equipment owners, particularly telephone companies and some government agencies. Manufacturers of larger systems have used several variations in system architecture to assure high levels of system reliability. In one configuration in which distributed stored program control is used, six microprocessors for line, trunk, resistor, system state, data base and console functions are employed. Each microprocessor and its associated memory are duplicated--one set is in the operational mode while

the other functions as a "hot standby." Another system configuration makes use of a dual control processor and a number of unduplicated microprocessor port controls. A failure of one of the latter means that the associated ports would be inoperative until the malfunction has been corrected; however, all other ports under the control of other operational microprocessors would still be functional.

Some equipment owners insist on batteries for use in the event of commercial power failure. Battery racks are available with or without earthquake bracing. In some lower cost systems, emergency power transfer consists of a relay that bypasses the switching system to connect one or more station lines directly to the serving central office.

As previously indicated, most modern systems are provided with elaborate software to perform maintenance diagnostics--to detect system faults, to assist the system in recovering from fault conditions and to initiate test routines for fault isolation. The software is designed for trouble report printouts. Access is through a data terminal which can be remotely located. In some systems, a running status of system "health" is provided by on-line diagnostics.

4.4.3 Specialty Switching Systems

Although stored program control TDM systems appear to be the preferred technology for the next two decades, some systems have appeared using other technologies. The Collins Model ATX-101 (recently discontinued), the Litton Amecom Terminal Communications System (TCSS)--installed at the Dallas-Ft. Worth Airport in 1975 for the FAA, and the 3M Brand CS^2 System II all make use of coaxial cable distribution. The Collins and the Litton systems employ stored program control FDM techniques.

The Collins systems could be configured up to 2000 stations; it made available wideband video channels, used dual processors to enhance system reliability, had a wide variety of operational features including maintenance diagnostics, and its installed costs were on the order of \$1200-\$1500 per station 31.

The Litton TCSS for the Dallas-Ft. Worth Airport was specially designed for FAA operations such as direct access intercom, with and without override; voice call; remote radio control; selective signaling; Automatic Terminal Information Service (ATIS); tape recording of all ATC voice communications; and access to and from long-line trunks. Additional requirements were imposed to allow for flexibility and growth for applications in the "Upgraded Third Generation" ATC system.

In the Litton system, switched voice and signaling are modulated (single-sideband AM) and then FDM multiplexed. Full duplex, equivalent 4-wire transmission is provided throughout the system. Multiplexed signals are transmitted in two bands, about 10 to 12.5 MHz and 12.5 to 15 MHz. With a channel spacing of 10 kHz, about 200 usable full duplex channels are obtained. Signaling is accomplished by a frequency shift keyed (FSK) signal located above the upper limit of the voice band (3000 Hz) in each FDM channel. Relatively low data transmission rates (300 and 600 baud) are used. A redundant 5 MHz reference signal is transmitted on the coaxial cable to all subsystems. To provide error detection capability, all signaling and data transmissions are encoded by the distributed control at each position and trunk. In response to each transmission error at a receiving unit, a retransmit request is returned to the transmitting unit which automatically reattempts the transmission and in this way the probability of lost calls is reduced. An uninterruptable power system is provided which is capable of operating the system for more than 30 minutes in the event the regular power source fails. The Litton system also provides for traffic data collection and assignment and modification of communications capability for each of the facility's ATC positions. If any position should fail, another position can be reconfigured to match the requirements of the failed position. An adjunct PBX provides access to the public switched network and the federal (FTS) network (32).

Although the 3M system* utilizes coaxial cable for distribution, it differs from the Collins and Litton systems in that it uses grouped pulse code modulation (PCM) or distributed phase modulation (PM) or a combination of the two. Common to all current system configurations is a coaxial cable spectrum divided into 42 6-Miz broadband channels--25 channels outgoing from the central office and 17 channels incoming to the central office. Each channel can transmit either analog or digital signals and also may be blocked and reused at remote terminals. The interface with analog central offices is a terminal having 24-channel PCM subscriber banks and broadbank equipment. (Subscriber banks are not required for interfacing a central office digital switch having 1.544 MBs or 64 KBs ports.)

In the distributed carrier configuration, the 3M system provides both voice and broadband signals directly to subscriber premises over a single coaxial cable. Each 6-MHz channel serves as one broadband channel or 24 250-kHz voice channels. Signals may be distributed to subscribers directly from a central office or from remote terminals. In this configuration, the system operates in a PCM mode from the central office to the remote terminal distribution unit, and from there to the subscribers' terminals in the PM mode. Power for the remote terminal is furnished over the coaxial center conductor.

---

*This system has been installed in several rural communities with the support of the Rural Electrification Agency (REA).

At the subscriber's premises, the coaxial cable terminates in a small box that separates voice and wideband signals for distribution to telephones to wideband terminals. It provides a 4-wire circuit to the telephone, eliminating the need for the hybrid network in the telephone set, and permits the use of an electronic ringer instead of the conventional electromechanical bell (33, 34).

Finally, and looking to the future, the potential for completely eliminating the need for premises wiring associated with even large PBXs exists through the development of wireless telephone systems utilizing low-power spread-spectrum modulation of an RF carrier which would afford complete privacy and permit a full complement of modern PBX operational features (35).

4.5 Forecast of Switching Machines

The toll telephone network has heretofore been predominately an analog voice network. While the local transmission network is largely digital, the emphasis there also has been on 4 KHz baseband analog service. Some auxiliary services have been provided on this network by indirect means, as for example low speed data by voiceband data modem. The military has provided secure voice communication on the public switched network by using sophisticated vocoder technology to obtain the relatively low data rates of voiceband data modems in conjunction with encryption equipment. Where higher bandwidth services are required, they have been provided on a

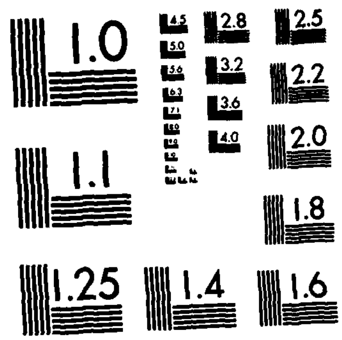
dedicated (non-switched) basis. Examples include television pictures and higher than voiceband data rates. Recently, the demand for switched data services has become sufficient to justify a proliferation of special-purpose data networks which have been set up independently of the switched public telephone network, such as the Xerox X-Ten and Satellite Business Systems (SBS). The Bell System has responded with a packet switched data network of its own in the Advanced Communications service (ACS).

The reason that these special services have been provided on a dedicated or overlayed switched network basis is that the bulk of the total traffic is still voiceband and adding to the expense of the public switched network to provide additional services could not be justified. The dedicated or overlayed network is a more economic approach as long as the special services remain a small fraction of the total traffic.

Several factors are emerging to radically change this picture. The special service offerings are increasing in volume at a much faster rate than the voice traffic, and this coupled with greater competitive pressures will lead to a provision of a wider range of services on the public switched network. Simultaneously, digital switching and transmission are now economically competitive even in a predominately voice network, and offer the capability for offering new services without an economic penalty in that voice network.

AD-A126 015 SOCIOECONOMIC IMPACT ASSESSMENT: COMMUNICATIONS INNUSTRY PHASE III TECHNO..(U) ACUMENICS RESEARCH AND TECHNOLOGY INC BETHESDA MD 02 FEB 79 FAA-APO-81-11-4 4/7

UNCLASSIFIED DOT-FA78WAI-932 F/G 17/2.1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

It is obvious, then, that the combination of digital transmission and switching, SPC switching, and CCIS, will be a very powerful force in leading to a wide range of new services on the public switched telephone network. While these technology factors enable provision of these services very economically, competitive pressures will motivate the administrators of the network (primarily the Bell System) to in fact provide them, presuming that regulatory constraints are not imposed.

There are many examples of new services which have been suggested (37-39). Among them are the following examples:

1. Various voice message services in which CCIS is combined with large voice digitization and storage facilities to enable customers to automatically forward short messages to other customers.

2. Teletext and viewdata systems in which home subscribers can access large data bases for up-to-date information on their home television screens.

3. Home surveillance systems in which various appliances are monitored and controlled by a central computer or distant subscriber and utility meters are read.

4. Electronic mail and magazine/newspaper distribution to homes and offices.

5. An automated office in which wordprocessing is combined with data base management and electronic mail to eliminate paper shuffling.

Some of these services can be provided by other means than over the telephone network. For example, CATV systems or even data transmitted during the blanking lines of the ordinary TV transmission can provide a data transmission capability into the home. However, these other networks have the distinct disadvantage of being distributional in nature, and (at least presently) not switched as is the telephone network. On the other hand, the telephone network has a lower bandwidth capability at present (about 80 kb/s for a digital subscriber loop). It appears that the only limitation this imposes is the inability to transmit TV pictures (except slow-scan or frame at a time).

With respect to office automation, most of the services described can be provided by computer vendors independent of the common carriers. However, there is a distinct advantage to providing them in the context of a PBX, which has the processor to perform the word processing functions, and existing telephone office wiring

to connect the processor to work stations, and the access to the nationwide telecommunications network to transmit the edited letters by electronic mail.

It is estimated that 70 billion pages of text are distributed among U.S. business and government organizations annually (33). Of these, 500 million are distributed by electronic means--facsimile, teleprinter, word processor, and remote batch printer. The need for better alternatives to moving paper around by traditional mail service is becoming obvious to all users. Statistical confirmation of this is shown by the annual growth rates of private delivery service (14%), mailgram (25%) and facsimile traffic (35%). While electronic document distribution is expanding, its potential has been scarcely tapped. At least two PBX manufacturers (ROLM and Northern Telecom) are offering electronic mail options with their new PBXs. A number of large organizations are reported to be designing and installing integrated electronic mail systems which include not only CRT electronic mail terminals, but communicating processors and facsimile terminals.

The trend toward new services which has just been described can be expected to modify the architecture of toll and local switching machines, in addition to PBXs, in the not too distant future (36). These changes will affect the switching network, the periphery of the switch, and the SPC processor.

The switch network, while remaining digital, will require the capability of providing a wide range of data rates for a given connection. Most likely, this will be achieved by dividing a high rate data stream into two or more standard rate data streams, transmitting them through the network, and then combining them on the other side by synchronous multiplexing. Thus, the primary modification would be at the boundary of the network, where this demultiplexing and multiplexing capability would have to be provided.

With respect to the SPC processor, much more processing capability will have to be provided. Two trends dictate this. First, many "calls" will be much shorter than they are today. For example, a single piece of electronic mail or page of facsimile will require only a very short connection time (less than a second). Thus, the processor will have to set-up and take-down a dramatically larger number of connections. Secondly, unlike today where the customer loses control of the switch once a connection is established, new services will require customer control of the switching machine during the course of a call.

Microprocessor technology will also make it more attractive to distribute the SPC processor, rather than centralizing it as in present designs. The processing capability can then be placed nearer to the points where it is needed.

Changes in the peripheral circuits will be required by the interface to a variety of different types of customer terminals and services. Thus, the peripheral circuits will contain processing capability to convert various protocols to a common protocol used within the switch. In addition, customer control of the switching machine during the call will necessitate that this processing capability be close to the customer interface.

Although it is speculative, one can also envision more radical changes in the switch architecture. The most likely such change is the provision of packet switching capability. The architecture we have described thus far is only capable of so-called circuit switching, in which a connection is established for the duration of a call. Circuit switching is inefficient for very short duration messages, because of all the resources consumed in setting up and

taking down the connection. In packet switching, short data messages from many sources and for many destinations are stored in a common memory and then forwarded over a shared transmission channel sequentially. This provision of storage and store-and-forward capability represents a rather large departure from present circuit-switched architectures.

In addition to the relatively glamorous trends we have described, the present evolution to larger and larger toll switching machines will continue, dictated by the growth of toll traffic and trunk-group efficiencies which result from larger machines. This evolution will, however, be relatively transparent to the end user.

The U.S. market for analog and digital central office switching equipment is expected to grow at a rate of 7% per year; however, starting from zero in 1975, the growth rate for digital switches is currently estimated at 45% per year with over $250 million of them sold by the end of 1979 (40). One half of the R&D for digital switches is the development of software for the smaller switches, and for the very large switches, software development costs can be three or four times the R&D costs for the hardware. Where years ago electromechanical central offices had 20 or more years of life

before obsolescence, the move to electronic designs will result in shorter 7-year design cycles, closely tracking the experience in the computer industry. (State and Federal regulators of the telephone industry are currently faced with substantially shortened depreciation schedules caused by technological obsolescence in this field--which has the effect of increasing telephone rates for all subscribers. This is at odds with their mandate to keep telephone rates at a reasonable level. This "balancing act" is a monumental task as billions of dollars in new telephone plants are proposed by the industry before the end of the century).

REFERENCES

1. Statistics of Communications Common Carriers, FCC, For year ended December 31, 1977, U.S. Government Printing Office, 1979.

2. Engineering and Operations in the Bell System, Bell Telephone Laboratories, Indianapolis, 1977.

3. "Technology of Telecommunications Switching," J.F. Day, Telecommunications Technology Symposium, IEEE-NTIA, May 22-23, 1979, Washington, D.C.

4. "Evolution of the SPC Network Plan," L.W. Helke, Proceedings of 1977 International Communications Conference.

5. "CCIS," Bell Laboratories Record, February, 1979.

6. "Stored Program Controlled Network," Bell Laboratories Record, March, 1979.

7. "An Introduction to No. 4 ESS," H.E. Vaughn, International Switching Symposium, 1972.

8. "No. 4 ESS," Bell Laboratories Record, December, 1977.

9. "Research Model for Time - Separation Integrated Communication," H.E. Vaughn, BSTJ, Vol. 38, July, 1959.

10. "New Developments in E-10 Digital Switching System," J. Bourbao, J.B. Jacob, International Switching Symposium, 1976.

11. "Experience Gained from the DCM Tandem and Local Exchange Field Trial Model (Moorgate)," D.J. Rose, B.A. Rayner, M.A. Henrion, International Switching Symposium, 1974.

12. "New Integrated Digital Switching System (IMA2)," J.R. Baichtal and J.C. McDonald, National Telecommunications Conference, 1976.

13. "No. 3 EAX System Description," R.K. Holm, GTE-Automatic Electric Journal, Vol. 15, July, 1977.

14. IEEE Transactions on Communications, "No. 4 ESS-Evolution of a Digital Switching System," R.A. Bruce, P.K. Giloth, E.H. Seigel, Vol. COM-27, July, 1979.

15. "Digital Switching - How It Has Developed," A.E. Joel, Jr., ibid.

16. "The LT-1 Connector - A Conventional Transmultiplexor," W.J. Mitchell, National Telecommunications Conference, 1978.

17. IEEE Transactions on Communications, "Synthesis of a T-Carrier Bit Stream Processing System: An Integrated Transmission and Switching System," J.C. McDonald and J.B. Baichtal, COM-26, May, 1978.

18. Internatioanl Conference on Communications, "Echo Suppressor Terminal for No. 4 ESS," R.C. Drechsler, June, 1976.

19. IEEE Transmittal Communications, "Synchronization for Telecommunications in a Switched Digital Network," C.A. Cooper, COM-27, July, 1979.

20. International Switching Symposium, "New Developments in E-10 Digital Switching System," J. Bourbao, J.B. Jacob, 1976, p. 421.

21. IEEE Spectrum, "Telephones Go Digital," S.G. Pitroda, October, 1979, p. 51.

22. Electronics, "All-Digital Telephone Instrument," October, 11, 1979, p. 80.

23. Engineering and Operations in the Bell System, Bell Telephone Laboratories, Indianapolis, Indiana, 1977.

24. Switching to Solid State Relays, Bell laboratories Record, February, 1978.

25. Com-Key® Telephone Systems, Bell Laboratories Record, March 19?

26. Horizon® Communications System, Bell Laboratories Record, November, 1977.

27. Sixty Years of PBX Development, Bell Laboratories Record, 1/6.

28. Principles of Switching, Paul Fleming, Vol. 10, Lee's Telephone Manuals, Geneva, Illinois, 1979.

29. PBX Systems Guide, Marketing Programs and Services Group, Gaithersburg, Maryland, 1979.

30. Showcase Seminar Focus on PABX, Digital CO and Fiber Optics Status, Telephony, July 16, 1979.

31. The Collins ATX-101 Audio-Video Communications System, H.A. Gutter, Business Communications Review, May-June, 1977.

32. Test and Evaluation of Dallas-Ft. Worth Terminal Communication Switching System, R.W. Cleary, FAA-NA-75-61, November, 1976.

33. Total Communications on Broadband Coax, Communications News, January, 1978.

34. Giving Telcos a Competitive Edge, Del Myers, Telephony, December 26, 1977.

35. Spread-Spectrum Techniques in a Wireless Telephone System, R.C. Dixon, 1977 Proceedings of the International Conference on Cybernetics and Society, Washington, D.C., September 19-21, 1977.

36. IEEE Transactions On Communication, "Digital Switching-Forces Which Shape Its Future," C.D. Hall, Vol. COM-27, July, 1979.

37. IEEE Spectrum, "The Wired Household," E.B. Carne, October, 1979, p. 61.

38. IEEE Spectrum, "Office Automation: A Challenge," N. Mokhoff, October, 1979, p. 66.

39. IEEE Communications Magazine, "New Services and their Introduction Into Existing Networks," W. A. Kaiser, July, 1979, p. 4.

40. Mavaricks Step Ahead of Giants in Digital Switch Race, Electronic News, August 7, 1978.

Chapter Five

Transmission

5. Transmission of Information in Telecommunication Systems

The third major ingredient of the telecommunications network is transmission, and if we seem poised at the threshold of substantial progress in the terminal and switching areas, progress in transmission is likely to be equally dramatic. The evolution of transmission was described in section 1. The first medium for the electronic transmission of information was wire pairs. With increasing demand for communication channels, these pairs soon became cables, and as diameters decreased to allow more pairs per cable, repeaters became required at decreasing distance intervals to amplify the signals to their original levels. Frequency-division multiplexing evolved to permit more than one signal to be carried over a single wire pair, and a continuous quest for broader-band transmission media and circuits began. Coaxial cables having a bandwidth approaching 10 MHz and a capacity of nearly 2000 simultaneous telephone channels were developed. For long-haul toll traffic, microwave radio became widely used after World War II, increasing transmission capacities and concentrating cost in repeaters having spacings of about 25 miles (dictated by the curvature of the earth). Geostationary satellite repeaters now allow these spacings to be circumvented, since one satellite can cover the entire continental United States.

The capacity and cost of such systems are largely determined by the satellite itself. The bandwidth of ground-based wire transmission has recently evolved further with the development of optical fibers having bandwidths of several hundred megahertz. Waveguides offer still another potential medium for high-capacity toll transmission.

As the type of information to be transmitted broadened from voice to include data and other forms of information, digital formats were developed and resulted in decreased error rates and lower cost. Several ways of more efficiently encoding analog signals in digital formats are being explored. Time-division multiplexing is becoming widespread, and bandwidth, once dedicated on a per-channel basis, is now being dynamically allocated as required by the information. As all of these new techniques have developed, they have not usually replaced the earlier approaches; rather they have supplemented them. Thus, today's transmission systems have available a broad range of technology and media: wire pairs, coaxial cables, microwave radio, fiberoptics, satellites, and wave-guides. For long-distance transmission, it is likely that microwave-satellite links and fiberoptics will be the dominant growth areas. New technology will be relatively easy to implement and network capacity is expected to increase rapidly. The local subscriber loops are a more difficult problem, however. With an increasing demand for two-way broadband communications on the part

of individual subscribers, the local loop provides the major bottleneck. These loops are dominated by wire pairs having limited bandwidth, represent an enormous investment in physical plant, and are unlikely to change significantly for several decades. Fiber-optics represents an attractive substitute for wire pairs in local loops but the cost of making the change is huge. Cable television-like systems using metallic or fiberoptic conductors is the goal but will be achieved gradually. Some combination of user-addressable microwave satellite broadcast system (for the down link) and a slower-speed ground link to a transmitting station is also a long-term possibility.

This chapter will discuss present activities in the transmission area. The first focus is on the efficient use of bandwidth through the development of efficient source coding, efficient modulation systems and through the allocation of the frequency spectrum. Emphasis in the discussion of spectrum allocation is placed on aeronautical communications, which are especially relevant to this study. The various transmission media are then discussed and forecasts are presented of their performance for the remainder of the twentieth century. Finally, some comments on the development of mobile telelphone and private communication networks are given.

5.1 Developments in Message Coding

The increasingly rapid trend to digital communications has many motivations. An increasing fraction of the information to be transmitted is in digital form, the processing and storing of data is easier in digital form, multiplexing is more easily implemented digitally, and digital modulation techniques are inherently more efficient from power and reliability standpoints. This is not to say that the transmission of analog data will decrease, but rather that when digital transmission techniques are employed, the analog data must be converted to a digital form. Indeed, one would expect that the demand for transmission of inherently analog data (voice and video) will increase correspondingly as the capabilities of digital communications rise.

This section reviews the state of the art in source coding for analog sources. This is the process of efficiently representing analog information in digital form. In this context, source coding is also called analog-to-digital conversion but is not restricted to linear binary codes as in section 2. The goals are to represent the data with as few bits per second as possible and to achieve high quality in the decoded reproduction. The lower the representation rate (in bits per second) the less channel capacity is required. Thus, an efficient representation often means a savings in power or bandwidth.

5.1.1 Types of Speech Coders

Roughly speaking there are two categories of speech coders: waveform coders and vocoders. Waveform coders attempt to accurately reproduce the speech waveform itself. On the other hand, vocoders merely attempt to produce a waveform that "sounds" like the original. Waveform coders include various forms of quantization, including adaptive and predictive forms. Pulse code modulation (PCM), differential pulse code modulation (DPCM), delta modulation (DM), and adaptive versions of these (APCM, ADPCM, ADM) are examples. Vocoders are more specifically tailored to speech. They include linear predictive coders and channel vocoders. There are other coders that do not fit neatly into either category. There are also some rather advanced waveform coders whose internal structure is more similar to the vocoders than the simpler waveform coders mentioned above. These include adaptive predictive codes (APC) and adaptive transform coders. Further, there are a variety of hybrids between waveform coders and vocoders.

The performance of a speech coder is determined primarily by the rate (in bits per second) used to represent speech information and the corresponding fidelity of the reproduction. The latter is difficult to quantify objectively. Therefore, fidelity must ordinarily be judged subjectively by human listeners for such

properties as intelligibility, speaker recognizability, and undesirable distortion. Other important factors when comparing speech encoders are robustness to speaker variations and environment, and robustness to errors in transmitting the encoded bits.

5.1.2 Encoding (Bit) Rate and Fidelity

Figure 5.1 (1) shows the general ranges of rates and fidelity for waveform coders and vocoders. The terms "toll", "commentary", "communications", and "synthetic" are used somewhat loosely to describe fidelity. "Toll" quality means comparable to standard telephone transmission of 3 KHz band-limited speech, "commentary" quality is closer to high fidelity, "communications" quality means highly intelligible but of noticeably poorer quality than toll, and "synthetic" quality means machine-like, sometimes with a loss of the original speaker's recognizable characteristics.

It is immediately apparent from Fig. 5.1 that there is a tradeoff between fidelity and bit rate. The waveform coders have higher fidelity and rate, while the vocoders have lower fidelity and rate. This is further borne out by the tables shown in Fig. 5.2 (1), which gives the bit rates required for various systems to achieve toll, communications, or synthetic fidelity.

DIGITAL CODING OF SPEECH

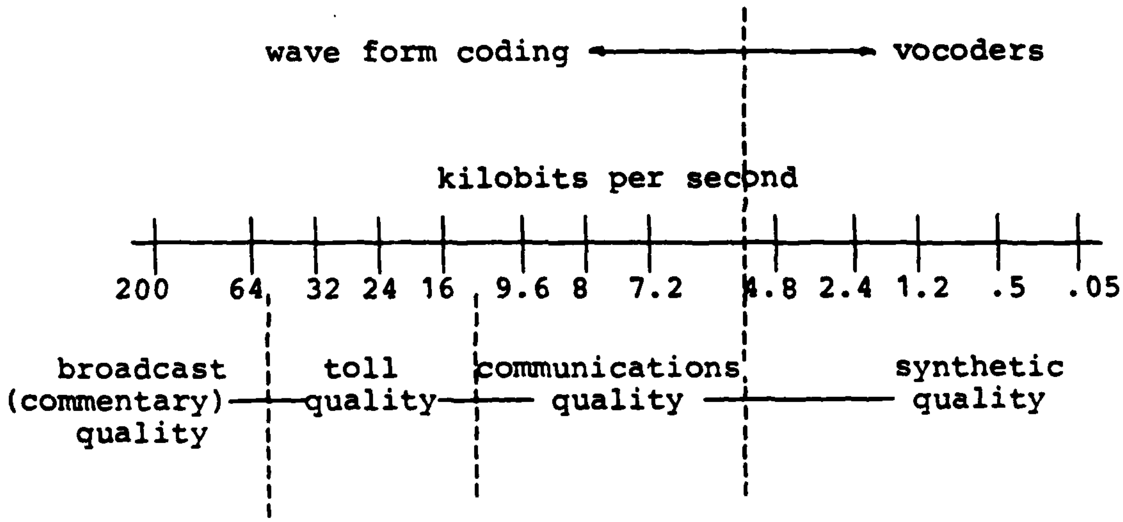


Fig. 5.1: The spectrum of speech coding transmission rates and their associated quality. (1)

Coder	kbits/s
Toll-Quality Transmission	
Log PCM	56
ADM	40
ADPCM	32
SUB-BAND	24
Pitch Predictive ADPCM	24
APC, ATC, ØV, VEV	16
Communications-Quality Transmission	
Log PCM	36
ADM	24
ADPCM	16
SUB-BAND	9.6
APC, ATC, ØV, VEV	7.2
Synthetic-Quality Transmission	
CV, LPC	2.4
ORTHOG	1.2
FORMANT	0.5

Fig. 5.2: Bit Rates Required for Various Coding Schemes and Fidelity Levels (1)

5.1.3 Robustness

Although waveform coders are generally fine-tuned to speech waveforms, they typically perform acceptably when used to transmit other sounds such as background noise, music, or the sound from two speakers. Vocoders generally do a poor job at transmitting anything other than the speech from a single speaker. Their performance when encoding a single speaker can be quite poor if there is significant background noise.

Most speech coders can tolerate transmission bit error rates as high as 10^{-3}. There are also robust versions of many coders that allow them to tolerate a bit error rate of 10^{-2}. In general, waveform coders tend to be less sensitive to transmission errors than vocoders unless precautions are taken.

5.1.4 Complexity and Cost

Figure 5.3 (1) gives the complexities (in corresponding numbers of logic gates) for various speech coders relative to adaptive delta and modulation (ADM). Roughly speaking, the hardware complexity increases as the bit rate decreases. In particular, the waveform coders are the simplest and the vocoders the most complex. Although cost is monotonic with complexity, using solid-state batch processing technology, it is by no means linear.

Relative Complexity†		Coder
1	ADM:	adaptive delta modulator
1	ADPCM:	adaptive differential PCM
5	SUB-BAND:	sub-band coder (with CCD filters)
5	P-P ADPCM:	pitch-predictive ADPCM
50	APC:	adaptive predictive coder
50	ATC:	adaptive transform coder
50	ΦV:	phase vocoder
50	VEV:	voice-excited vocoder
100	LPC:	linear-predictive coefficient (vocoder)
100	CV:	channel vocoder
200	ORTHOG:	LPC vocoder with orthogonalized coefficients
500	FORMANT:	formant vocoder
1000	ARTICULATORY:	vocal-tract synthesizer; synthesis from printed English text.

† Essentially a relative count of logic gates. These numbers are very approximate, and depend upon circuit architecture. By way of comparison, Log PCM falls in the range 1-5.

Fig. 5.3: Complexity of Various Voice-Coding Approaches

5.1.5 Present and Future Coding Activity

The Bell system now uses 64 Kbps log PCM, described previously, on their digital transmission lines. This coding format requires a relatively high bit rate but is capable of toll quality, is relatively simple to implement in hardware, and is capable of encoding signals other than speech. Many of the encoders described in this report have been built in prototype hardware form. Some are in production, particularly for military applications. For example, continuously-variable-slope delta modulation (CVSD) at 16 and 32 Kbps has been implemented on a single chip, and LPC coders have been implemented on microprocessors and on special-purpose programmable processors.

It is expected that waveform coders are ready for large scale production and wide application with present day technology. Concerning vocoders, a 1978 report (2) said, "Given the present-day and near-future advances in technology, it is predictable that present-day vocoder algorithms will soon be implementable as low cost compact devices. How soon this comes to pass does not depend on any new technological breakthroughs but rather on the direction of investments in LSI development realizable with present device technology." As the rapid growth of VLSI technology continues, the cost of source coding will decrease and more complex, efficient coding schemes will become attractive. For communications and toll

quality, the required bit rates could decrease by factors of at least four or five during the next 20 years, with a corresponding increase in the capacity of existing channel bandwidths.

5.1.6 Digital Speech Interpolation

A technique for obtaining a reduction in the bit rate required for the transmission of speech is Digital Speech Interpolation (DSI). This technique is closely related to the Time-Assignment Speech Interpolation (TASI) which has been used on submarine cable transmission since 1960. Both TASI and DSI take advantage of the fact that in conversational speech only one direction of transmission carries active speech at one time, and furthermore, even in the active direction there are pauses in the speech between phrases, words, and sentences. Therefore, there is need to transmit each trunk only about 30% to 40% of the time. TASI and DSI take a large number of trunks and compress them to a smaller number of channels for transmission by using the channels to transmit only the active trunks. The number of channels can then be adjusted to make it statistically improbable that the number of active trunks exceeds the number of available channels.

TASI and DSI differ primarily in that TASI is best suited to analog transmission and DSI is best suited to digital transmission. In TASI, when the number of active trunks exceeds the number of channels, some trunks are simply frozen out (not transmitted).

This is subjectively very undesirable, so the probability of this happening must be kept very small, limiting the compression ratio (ratio of trunks to channels). In DSI, advantage is taken of the fact that the bit rate of each transmitted trunk can be adjusted to the point that all active trunks can be accomodated on the transmission facility regardless of the number of active trunks (3, 4). Thus, the freeze-out degradation of TASI is replaced by the more palatable impairment of increased quantization distortion. As a consequence, the compression ratio can be set somewhat higher in DSI as compared to TASI.

An additional feature of DSI is that the bit rate reduction required to compress all the active trunks into the available channel capacity can be done more efficiently using any of the waveform coding techniques described earlier in this section. The bottom line is that in DSI a comparable quality to that of 64 kb/s speech can be obtained at an equivalent bit rate of approximately 16 kb/s.

The obvious application of DSI is in common-carrier transmission (where large numbers of voice trunks are available simultaneously) and in long-haul applications, where transmission costs dominate. (The extra terminal costs of DSI cannot be justified in short-haul applications.) Since the predominance of digital transmission is

presently short-haul, DSI has not made an impact. However, DSI is an attractive means of making long-haul digital transmission economical, and will undoubtedly see extensive application within the next decade.

5.2 Digital Modulation Techniques

As mentioned in other parts of this report, there is an increasing trend towards and demand for digital communications. This necessitates the development of more efficient modulation techniques because the key resources of power and bandwidth will only become more costly to use. In this section we discuss digital modulation techniques that are currently being employed, those that are under development, and those that might be employed in the future.

From the user's "end to end" point of view the primary characteristics of a digital communication system are the bit rate and the error rate. The former is the message throughput rate in bits per second; the latter is the fraction of these bits that are in error. Such systems use two costly resources: power and bandwidth. An efficient system is one that achieves large bit rate and low error rate with a minimum of power and bandwidth. The overall efficiency of a communication system is determined by several components: the modulation system (modulator-demodulator), the transmitter, the antennas (transmitting and receiving), the

propagation medium, and the receiver. By and large these components can be studied separately. In this section, we discuss the efficiency of various modulation systems. It is to be expected that any improvements therein will cause a corresponding improvement in the overall efficiency.

It is customary to evaluate and compare the performances of modulation systems in the presence of white Gaussian noise. Such noise is always present. While it may not always be the predominant noise, comparisons based on this assumption generally hold up under actual operating conditions. In this environment the efficiency of any type of digital modulation system (e.g., phase shift keying or frequency shift keying) can be conveniently characterized by two numbers: the bandwidth efficiency b and the energy efficiency e. The former is the ratio of the rate R to the transmission bandwidth W. For purposes of comparison the 3 db bandwidth will be used. The latter is the ratio E_v/N_0 of the received signal energy per information bit to the (one-sided) power density required to achieve the nominal error rate of 10^{-4}. In these terms a good system has large b and small e. Of course, the complexity of implementation is another fundamental cost determining factor. In general there is a tradeoff due to the fact that the modulation systems with better efficiencies are more costly to build. Hence the systems with better efficiencies are not used unless the savings

in the resources of power and bandwidth justify the increased cost of implementation. In the future one should expect power and bandwidth to become more expensive, hence the more sophisticated modulation techniques will be used.

The two parameters b and e have been plotted in Figure 5.4 as points (b,e) for a variety of modulation techniques. These points represent idealized performance. In actual systems b comes within 1 or 2 db of the given values and e is within 10-20%. Also plotted is the curve of optimal efficiency derived from Shannon's channel capacity formula. No modulation system with very low error rate can have its (b,e) point lie below this curve.

For the purposes of discussion we isolate three regions in Figure 5.4. The first is the central region that includes noncoherent binary frequency shift keying (NFSK), noncoherent on-off keying (NOOK), differential phase shift keying (DPSK), binary phase shift keying (PSK), and quadrature phase shift keying (QPSK). These are the standard modulation types used widely today. They are listed in order of increasing efficiency and increasing complexity of implementation. Generally speaking, region 1 contains the least complex modulation techniques. They have energy efficiencies ranging from 8.4 to 12.5 db and bandwidth efficiences ranging from about 0.7 to 2. There are other modulation systems that are not

ACUMENICS

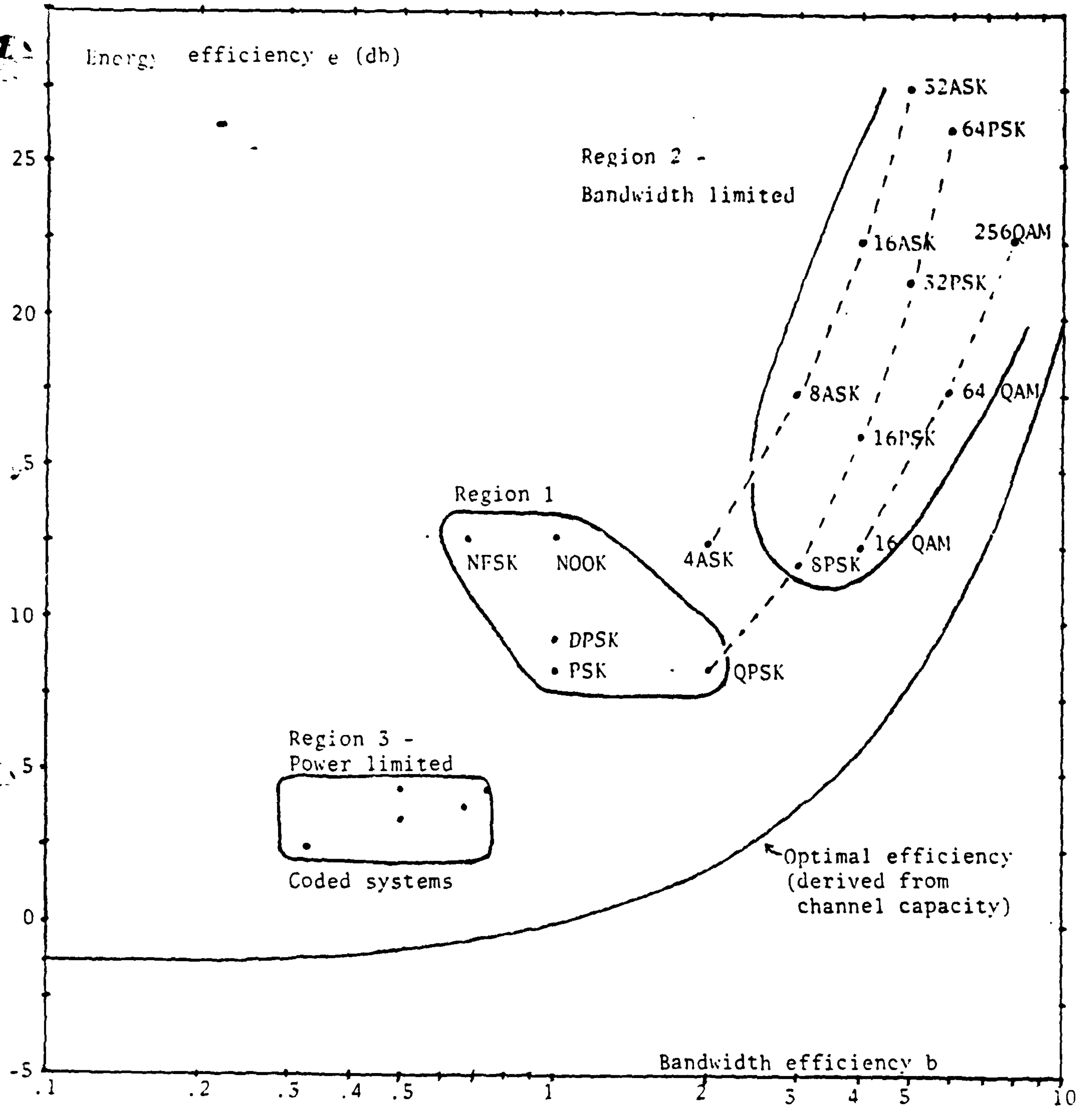


Figure 5-4: Energy efficiency e (in db) vs. bandwidth efficiency b. $e = E_b/N_o$ needed for error rate 10^{-4}. $b = R/W$.

ACUMENICS

shown that also fall in region 1, such as offset quaternary phase shift keying (OQPSK), minimum shift keying (MSK), and various partial response systems. These systems are especially adapted to certain specific problems; for example, MSK is adapted to the problem of nonlinear amplification in satellite repeaters.

In many communication environments bandwidth is in short supply. Consequently, when there is demand for higher bit rates, it can only be met by using a bandwidth efficient modulation technique like those in region 2 of Figure 5.4, the so-called bandwidth limited region. These include forms of multi-level amplitude shift keying (e.g. 8-ASK), phase shift keying (e.g. 8-PSK), and quadrature amplitude modulation (e.g. 16 QAM).

Wire communications have been operating in the bandwidth limited region for some time. Hence, there has been considerable effort devoted to bandwidth efficient modems. At the present time, microwave radio below 20 MHz in both ground and satellite based systems is power limited. Consequently, there is considerable effort presently underway by the common carriers and the military to implement systems with bandwidth efficiences of 3 to 4. These will be widespread by the year 2000. If the demand for bit rate increases still further, then development of systems with even higher bandwidth efficiencies will be necessary. As can be seen from Figure 5.4, in theory such systems can be built.

When, on the other hand, power is in short supply, then the coded communication systems of region 3, the power limited region, are called for. These have energy efficiencies that are from 3 to 6 db better than those of regions 1 and 2, but have 1.3 to 4 times less bandwidth efficiency. Furthermore, these techniques tend to be more complex and, hence, costly. However, they are constructed primarily of digital hardware, and their cost can be expected to follow the generally decreasing trend of such devices. Coded systems (e.g., convolutionally coded systems) have been under continual development for a long time at a low level. One should expect this to continue, and as it does one should expect to find an increasing number of communication problems in which coding is a cost effective solution. Although bandwidth limited environments are predominant, there are important power limited environments where coding might be the best solution. For example, they have been frequently used in space communication, and might prove very effective for low power emergency or telemetry communications. Indeed, the extreme shortage of bandwidth in wire and microwave radio has promoted the development of other communications media that are very broadband and hence, utilimately, power limited (e.f., microwave radio above 20 MHz and optical fibers). In the upper microwave region, a predominant effect is fading, which is effectively a low power phenomenon. Coding is likely to be effective here. In optical fibers coding might allow increased spacing between repeaters.

In conclusion, we anticipate that the ever increasing demand for digital communications will stimulate the design of more efficient modulation systems such as those in regions 2 and 3 of Figure 5.4.

5.3 Spectrum Allocation and Utilization

The radio spectrum is a limited natural resource and like most resources is susceptible to oversubscription and pollution. The demand for radio frequency channels, particularly in the U.S., is skyrocketing as the use of established services increases and as new services are established. Some idea of the current and future demand is shown by the following estimates:

- The number of FCC licenses for special and safety transmitters (amateurs, CB, land mobile, aviation and marine) is estimated to be about 25 million for 1980 and has been growing at the rate of about 8 percent per year.
- The number of aircraft (air carrier, commuter, and general aviation) is expected to increase from 191,000 in 1980 to over 470,000 in the year 2000.
- The number of VFR and IFR operations (takeoffs, landings, or calls to an FAA ARTC Center) is expected to increase from about 175 million in 1980 to about 365 million in 2000.
- The number of avionic equipment installations (communications voice and data systems, ELT, ATC transponders, DME, radio altimeters, Dopoler systems, HF, UHF public telephone and weather radar) is expected to triple -- from about 800,000 installations for general aviation, air carriers and the military in 1980 to about 2.4 million in the year 2000. All require radio frequencies for their operation (10).

5.3.1 The Allocation Process

The first use of radio communication was in 1896 between ships and shore after Marconi demonstrated his development of wireless telegraphy. Soon thereafter, in 1901, his signals bridged the Atlantic and a whole new era of instant worldwide communications supplemented the existing underseas cable and landline telephone and telegraph systems.

The first allocation of frequencies was made at the Berlin Radio Conference in 1906, which designated 500 and 1000 KHz for public service frequencies for ship-to-shore telegraphy. The Berlin Radio Conference was convened because the British Marconi Company in 1902 refused to deal with its German competition in relaying a message from Prince Henry of Prussia to President Theodore Roosevelt. In 1912 the Internation Radio Conference was held in London where spectrum allocations were made for the 150 KHz to 1000 KHz band. Since that time over 60 different radio services have been given allocations from 10 KHz to 275 KHz at a number of World Administrative Radio Conferences held under the sponsorship of the International Telecommunications Union (ITU).*

Because radio signals do not recognize national boundaries, frequency allocation is subject to both international and national regulation. International and national requirements have led to

*The International Telegraph Union was formed in 1865; the International Radio Telegraph Union in 1906; and the two were merged in 1932.

the apportionment of specific radio frequency bands for use by specified radio services. Such apportionments are known as frequency allocations and are differentiated from "frequency assignments" which are authorizations for the use of discrete frequencies by specific radio stations.

The nations of the world have found it necessary to work together to prevent chaos in international communications and have found a method through the ITU, which has been responsible for a long series of international conferences for this purpose. Among its other responsibilities, the ITU functions to allocate the radio frequency spectrum in order to avoid harmful interference* between radio stations in different countries and to improve the use made of the radio spectrum.

Within the United States, frequency allocations are managed by the Federal Communications Commission (FCC) and the National Telecommunications and Information Administration (NTIA). The Communications Act of 1934 vests in the FCC the responsibility for the regulation of non-Government interstate and foreign telecommunications. The Act, recognizing the Constitutional powers of the President, exempts radio stations belonging to and operated by the Government from the Act and provides that such stations shall use

---

*Reports of radio frequency interference were made as early as 1907. Scientific American, January 1907, reported, "The lack of transmission selectivity has brought about a state of affairs that borders on chaos; for only one or two stations in the active zone--and this means a radius of 1000 miles--can send at one time."

such frequencies as shall be assigned by the President. The Interdepartmental Radio Advisor Committee (IRAC) was organized by Herbert Hoover in 1922, who was then the Secretary of Commerce and responsible for administering the Radio Act of 1912. IRAC assumed the responsibility for the President in coordinating and assigning frequencies for the use of Government stations, including those of the Federal Aviation Administration (FAA), and currently functions under NTIA.

At the international level, frequency allocations are made regionally. Region 1 includes Europe, Africa, and Northern Asia. Region 2 includes North and South America and the Central Pacific, and region 3 encompasses Southern Asia, Australia and the Southern Pacific.

In recognition of the dual jurisdiction over the spectrum in the U.S. (by the FCC for non-Government services and NTIA for Government services), it has been divided into three categories: exclusive non-Government, exclusive Government and bands shared by non-Government and Government users.

The NTIA Table of Frequency Allocations (11) lists about 70 Government and non-Government aviation-related frequency band allocations between 10 KHz and 265 GHz. Figure 5.5 provides an overview of radio frequency allocations for various Government and non-Government aviation requirements.

Frequency Band (MHz)	Application	Frequency Band (MHz)	Application
0.01-0.014	Omega Radionav.	2700-2900	Ground Radars
0.09-0.110	Loran C Radionav.	3500-3700	Ground Radars
0.11-0.130	Aircraft Comm.	4200-4400	Radar Altimeters
0.16-0.490	Radionavigation	5000-5250	MLS
0.51-0.525	Radionavigation	5350-5470	Airborne Weather Radar
1.605-1.800	Loran A	8750-8850	Doppler Radar
2.85-22.00	20 HF Bands (Long Dist. Flying)	9000-9200	Aircraft Beacons Approach Radar
74.6-75.4	Aero Marker Beacons	9300-9500	Airborne Weather Radar
108-136	Air Traffic Control En Route Stations Search and Rescue Flight Tests	13250-13400	Airborne Doppler Radar
225-328.6	ATC Military	14000-14300	Sweep Radar, Small Aircraft
328.6-335.4	ILS Glide Path	15400-17700	MLS, Airborne Weather Radar
454.7-459.95	Public Correspond.	31800-33400	Sweep Radar, Small Aircraft
960-1215	TACAN, DME, MLS, ATCRBS	43000-265000	At least six satellite-aircra[ft] radionavigation allocations
1215-1300	Ground-based Radar		
1300-1350	Long-range Radar		
1525-1660	Aeron. Telemetry		
1535-1660	Satellite Aircraft Collision Avoidance		

Figure 5-5: Radio Frequency Spectrum Allocations for Aeronautical Communications

5.3.2 The Frequency Assignment Process

The International Radio Regulations of the ITU and the Rules of the FCC provide for the allocation of frequencies for use by specific radio services, the allotment of segments of those bands for use in specific geographical areas, and the assignment of specific frequencies in these bands to individual radio stations.

In the assignment of frequencies, there are a number of factors to be considered in order to establish protective criteria for individual services. Some of the criteria involve safety of life and property, the uniqueness of the needs of the user and the susceptibility of that service to interference. As a consequence, there are services that require a high degree of protection, some a moderate degree, and some which receive virtually no protection.

Among the highly protected services are aeronautical and maritime services. The public also expects interference-free service from its AM, FM and TV broadcasts. In addition, common-carrier communication satellites, high seas and VHF aeronautical and maritime public correspondence, and domestic land mobile systems require considerable protection. Anomolies are the radio astronomy and standard frequency services which are protected but with no connotation of safety to life or property.

Services receiving moderate protection include police, fire, highway maintenance, forestry and public utility services, as examples. Those who receive practically no protection include the amateur, CB, and business radio services.

In granting station licenses, the FCC controls transmitter characteristics (or emission sources). While the Communications Act of 1934 was silent on providing controls on receiver selectivity, the Commission does control radiation from receivers. The Commission's Rules place requirements on such characteristics as transmitter power output, frequency stability, types of emission (methods of modulation), bandwidth of emission and in some instances the hours of operation. The FCC establishes minimum standards with respect to frequency stablity, freedom from spurious emissions and the like and specifies that the authority to operate a device will be granted only if it has been type-approved or type-accepted. Type approval is limited to a determination that the equipment is capable of complying with the technical requirements of the rules under which it will be operated, so long as the equipment is not modified, and assuming that it will be operated and maintained properly. Type acceptance is based on representations and test data submited by the manufacturer or prospective licensee, and again, is limited to a determination that the equipment appears capable of meeting the technical requirements if properly operated and maintained.

A wide variety of devices come under FCC jurisdiction, including incidental radiation devices, restricted radiation devices, and industrial/scientific/medical devices. An incidental radiation device is one that radiates RF energy when in operation although it was not designed to do so. Typical in this class are fluorescent lights, automobile ignition systems, electric motors and thermostats. Restricted radiation devices are those which generate RF energy intentionally, but are not candidates for station licenses nor fall under the rules dealing with industrial, scientific and medical equipment. This class includes garage door openers, wireless microphones, cordless telephones, and radio and TV receivers. These devices can present problems. One example resulted when a high-powered military transmitter was reactived on 150 MHz several years ago causing garage doors over a seven state area to go berserk everytime the transmitter went on the air. Another example is a telephone answering machine that, when activated, broadcasted all of its recorded messages in the FM band over an area of several blocks.

Industrial, scientific and medical (ISM) devices are covered in a separate section of the FCC rules and include medical diathermy equipment, industrial heating equipment used in production processes, and equipment which applies RF energy to materials to produce physical, biological or chemical changes. The rules limit harmonics and spurious emissions from such devices to protect other authorized radio services.

In aviation, there are aircraft that fly specific routes (generally civil) and those that fly off-route (generally military). The international allocation tables designate "R" services for civil users and "OR" services for military. The "OR" allotment plan simply designates specific frequencies for specific countries; but the "R" plan is more complicated and is based on a formula devised at the Atlantic City Radio Conference in 1947, using as load factors the actual number of commercial aircraft scheduled to fly between points A and B on a specific route and on a predetermined date. From this, routes were divided into two basic categories--Major World Air Route Areas, and Regional and Domestic Air Route Areas. Families of frequencies were then selected from the several bands allocated to the service so as to permit communications at any time between aircraft in flight and ground stations. Ground stations were then separated sufficiently to ensure a 15 db ratio of desired-to-undesired signal at the limit of the service range of the ground station. A 20 db co-channel separation is now required.

At the present time, the U.S. uses HF (R) frequencies only in international aeronautical service which are accommodated in a number of bands between 2850 to 22,000 KHz. Domestically, civil air traffic control stations are operated by the FAA on VHF between

117.975 and 136.0 MHz in 506 25 KHz channels. Double side band amplitude modulation is standard. Within this band, a suballocation (128.825 to 132.0 MHz) provides for enroute stations owned by non-Government licensees to communicate for business purposes with air-carrier aircraft. These are licensed by the FCC, and are owned and operated by Aeronautical Radio, Inc. (ARINC) for most of the U.S. domestic airlines.

The communications common carriers, in applying for microwave system licenses, make extensive engineering analyses of the design of such systems, the use of frequencies, and potential interference between these systems and communication satellite systems. Analysis of a single potential interference situation may require 300 man-hours of effort or more. Computer programs have been developed to facilitate the analysis of such situations prior to the submission of the application to the FCC.

In 1978, there were a total of 24,486,139 FCC-licensed transmitters, including 21,000,000 in CB service, 13,362 AM, FM and TV stations, and 195,411 fixed and mobile aeronautical stations (with 734,637 transmitters) (12). There were a total of 4,137,137 licensed commercial radio operators.

Government spectrum usage is subjected to more policy consideration, coordination and engineering than is generally known. The process

begins with the assignment of a mission by Congress to a Government agency.. Upon approval and appropriation of funds, the Government agency determines its communications requirements. The agency frequency manager examines the frequency allocations and frequencies available to the agency. When it is evident that the required frequencies are not available, IRAC/NTIA is consulted. When new types of communications-electronics systems are contemplated, an electromagnetic compatibility analysis is made. Where frequency support is considered possible without any electromagnetic incompatibility, the agency frequency manager selects possible frequencies, makes any required engineering evaluations, coordinates the selection with other agencies involved, and files an application with IRAC. The IRAC Secretariat places the application on the agenda of the Frequency Assignment Subcommittee for distribution to each agency and the FCC for study. Over 50,000 applications are processed annually by the IRAC Secretariat (9). In 1975, the military purchased more than $10 billion in electronic systems and had about 50 percent of all Government frequency assignments; whereas the FAA had about 15 percent--or close to 20,000 frequency assignments.

5.3.3 Present Concerns Relative to Aeronautical Spectrum Utilization

In September 1979 a World Administrative Radio Conference (WARC) sponsored by the ITU convened in Geneva for the purpose of revising the International Radio Regulations. Over 1200 delegates from 154 nations participated. This number of countries is almost double the 84 attending a similar meeting in 1959. Scores of developing nations are concerned that there will not be adequate radio frequencies or satellite positions available by the time they become technologically advanced to use them for their own domestic needs. Some equitorial countries, particularly Columbia and Ecuador, also proposed that they be given air rights over communication satellites positioned or passing over their territory.

From a review of the suggested changes in international regulations made in preparation for the WARC (13, 14), three major present issues involving aviation services emerge:

1) Air-ground communcations are expected to at least double within the next twenty years and are expected to place a severe demand on the availability of radio channels (15).

The Aviation Service Working Group, responsible for preparing the U.S. position for the 1979-WARC, felt that the introduction of improved frequency utilization could best be accomplished by conducting studies and experiments at 136-138 MHz, and then

introducing new techniques into 118-136 MHz, if findings support such action. The FCC essentially agreed. Some techniques might include splitting the current 25 KHz channels, but maintaining the current DSB AM modulation, or even employing 3 KHz SSB modulation.

2) Air-ground public correspondence. The Radio Technical Commission for Aeronautics (RTCA) (14) states, "As the use of General Aviation airplanes grows in the United States, the need for a communications link between the executive in the airplane and his business requirements on the ground becomes more and more pressing."

In the U.S., public air-ground radiotelephone service was authorized by the FCC in December 1969, and consisted of twelve 2-way channels in the 454.7-459.95 MHz band. A thirteenth up-call receive-only channel provides for a ground station to ring any given aircraft telephone. A five digit system provides for over 50,000 different telephone numbers to be selectively called. An aircraft at 10,000 feet has a range of 150 miles, at 40,000 feet the range is 300 miles. There are 92 planned sites in the U.S. having from one to four channels each and in 1976 there were 29 stations in operation serving about 1000 sets in General Aviation aircraft. The airlines have not participated. For the airlines to participate, RTCA estimated that at least 60 channel pairs would be required, although there have been other estimates indicating that more than 200 channel pairs might be necessary. Although requested by RTCA, the FCC was not persuaded that sufficient need

had been demonstrated to justify an exclusive high frequency spectrum for public correspondence communications with aircraft other than those of the scheduled air carriers.

3) Satellite facilities for aeronautical services.

Despite a pessimistic near-term outlook for an operational system, the FCC has noted the aeronautical community's concern that existing frequency allocations may be insufficient to meet planning purposes. Additionally, the Commission has recognized potential needs at VHF, and at the 1535-1660 MHz and 5000-5250 MHz bands. In preparation for the now defunct AEROSAT program, Boeing 747 aircraft were fitted with VHF satellite antennas and limited satellite service within the present VHF band could be implemented in a short time, should the need develop.

The 1535-1660 MHz band is structured to make provision for the Global Positioning Satellite (GPS) System at 1575 $\pm$ 12 MHz. The Commission noted that aeronautical communications satellites and GPS can successfully share spectrum (GPS uses spread-spectrum technology). Therefore, these allocations should permit a common GPS/communications package to be placed on the same aircraft and provide for common receiver design. Also, the 5000-5250 MHz band has always been intended for links between land stations and satellites, and the U.S. proposal would permit the connection of ATC centers via satellites.

5.3.4 Spectrum Engineering

Spectrum engineering has become an exceedingly complex subject involving frequency selection, engineering design considerations and interference reduction to meet the needs of a growing army of users whose interest span narrow-band and wide-band applications and short-distance and long-distance communications requirements. In a word, spectrum engineering has as its prime purpose the optimization of spectrum resources (16).

Looking to the future, there appears to be several thrusts developing that have the potential for coping with the explosion in new requirements and uses:

1) Making more frequencies available.

Frequencies up to 10 GHz are literally filled to the brim with users. Enormous pressures are always present in which one service attempts to rationalize its position for encroaching on the allocation of another service. And, as some services are retired, readjustments are made from time to time. An example is the phasing out of Loran A starting in 1978.

A relatively untapped reservoir is the band from 10-300 GHz. Although it appears in Fig. 5.5, two major problems associated with its use are propagation losses (associated with precipitation, water vapor, and oxygen absorption), and relatively high system development costs. As an example, when considering communications links

operating in the 10-40 GHz band, rain attenuation is the dominant consideration. For high-availability communications links (on the order of 0.5% outage or less), the typical distance per hop for a terrestrial relay varies from 1 to 10 km at 10 GHz to about 0.2 km at 40 GHz.

Offsetting the losses from rain attenuation as frequency increases, the available bandwidth per link also increases, suggesting that operation near 10 GHz is better suited for applications requiring higher availability and lower data rates. Higher frequencies (nearer 50 GHz) are better suited for short range links with lower availability requirements and wideband or high data rate transmissions.

All RF components, including low-noise amplifiers, RF power sources and mixers, exhibit increasing costs as the frequency jumps from 10 to 40 GHz due to limitations in available technology and more difficult fabrication techniques. Oscillator stability, along with increasing frequencies, requires a mix of technologies and additional costs. (One can expect 0.3% stability from free-running oscillators and 0.001% stability from crystal-referenced sources.) Also, more efficient modulation and encoding techniques require higher performance and more expensive transmission equipment by

requiring more linearity in equipment design to offset greater losses due to channel and equipment impairments. It has also been found in regard to interference between digital and analog carriers in adjacent slots that analog carriers show greater sensitivity to interference than adjacent digital carriers or adjacent analog carriers (17).

2) Channel Splitting

Channel splitting has been used many times in the VHF aviation band--from 100 KHz to 50 KHz and now to 25 KHz. Most likely it will be split again before the end of the century. The end result is more channels through hardware improvement (better selectivity). The change to 25 KHz fortuitously came at the time of the introduction of solid-state technology which resulted not only in the required channel width, but also in equipment that is smaller, lighter, more reliable, has more features and requires less power to do the same job.

3) Operational Fixes

Most aviation services are safety oriented and, as such, they are less amenable to operational fixes than are other services, many of which can make use of time sharing, geographical sharing, sector blanking, reduced power levels, local coordination, and the use of special hardware (e.g., sets that can tune to only one or a few frequencies).

4) Alternatives to the RF Spectrum

It is possible that the development of alternative wideband transmission media could relieve some of the anticipated future demand for the RF spectrum and could be more cost effective than the implementation of specialized radio channels. As examples, communications common carriers now offer a variety of services for both narrow band and wideband voice and data channels. Some CATV operators are leasing unused channels for data transmission. Finally, for applications requiring short distances and medium data rates, a cost-effective solution appears to be free-space digital optical communications links making use of laser technology, microwave systems. One recent system uses a laser diode, an avalanche photodiode and PPM modulation. It has a 2.7 km range, a 6.3 Mbs data rate and a 10^{-6} error rate (18).

5) Improving Spectrum Utilization Through New Concepts

Two promising developments are the use of digital technology for air-ground communications and the recent experiments with spread-spectrum concepts, both of which will be reviewed here as both have the potential for improved spectrum utilization.

The RTCA Special Committee 120 was set up to prepare a proposal for the U.S. National Aviation Standard for VHF/UHF air-ground communications (voice and data) (24). In the scenario presented, it was felt that the transition to the air-ground communications system of the 1980s will need to proceed in a manner that will

coordinate with the increasing automation of the ATC system. The attractiveness of the digital mode of communication is that it is the natural means of communication with computers and that the information transfer rate (at 4800 bps) is considerably greater than normal voice communication between people.

Automation of the enroute and terminal air traffic control systems has commenced with the computer processing of plan position, aircraft identity, and altitude data (derived from the radar and radar beacon surveillance system) to provide controllers with information on which to base their decision making. The next phases will selectively transfer to the computer the tasks of 1) organizing traffic flow for optimum utilization of airspace and terminal facilities, 2) the generation management and delivery of other ATC supporting control messages, and 3) determining how predicted conflict information can best be handled. The FAA is currently undertaking a wide range of experimental activities for the introduction of automated ATC processes and supporting air-ground communications, including arrival runway metering and spacing; frequency changes, beacon code assignments and altimeter setting delivery; conflict detection and resolution including provisions for terrain and controlled and restricted airspace avoidance; automatic monitoring of simultaneous parallel approaches to parallel runways; automatic wake vortex monitoring; advanced metering and

spacing including provisions for departure and arrival sequencing and multiple runway as well as multiple airport traffic sequencing; Automated Terminal Information Service (ATIS) delivery; and airport pre-departure information including "pushback" clearances, ATC enroute clearance delivery and airport surface movement guidance and control.

During the introductory phases of ATC control message automation, it is anticipated that controllers will, as a minimum, be called upon to monitor many of the computer-generated automatic processes. As experience with the computer's performance accumulates and confidence is gained in its capability, the need for human monitoring or supervision of automatic processes will diminish.

The earliest automated computer-generated ATC air-ground communications mode envisioned could utilize computer-generated voice response to assist in system monitoring. Subsequent user acceptance of the automatic ATC communications concept can be expected to encourage increasing use of visual display/synthetic voice readouts in the aircraft fleet on an evolutionary basis. It may be anticipated that in areas of low traffic density, parts of the system would not be utilized.

Phase I of the RTCA scenario envisions direct ATC computer-to-aircraft communications for instructions generated by early capabilities of the FAA's Upgraded Third Generation ATC system.

Voice exchanges between human beings would be used to resolve any problems. Phase II of the scenario would follow the extension of the ATC system automation to encompass the compilation and issuance of all types of clearances. The design of the automation capability for provision of information services such as ATIS would permit outputs to suitably equipped aircraft without human intervention on the ground.

The second area to be considered relates to spread-spectrum techniques (19, 20). These are well adapted for use in civil communications, although they have been confined almost exclusively to military systems. The spread-spectrum concept is based on the premise that a given frequency space can be shared among several users simultaneously, provided the respective carriers are randomly varied with respect to each other so as to remain unsynchronized. The capabilities of spread-spectrum techniques are listed below.

Capability	Application and Remarks
Multiple access	Provision for many users of a given spectrum space. (Time division and/or code division access are readily implemented.)
Discrete addressing	Provision for selection of a specific receiver. (Receivers are addressed by code selection.)
Low power density	Reduction of interference to conventional systems. (Spread-spectrum signals distribute their power over wide time-frequency area.)

Interference rejection	Reduction of the effect of co-channel users. (Receivers minimally affected by conventional signals on the same channel.)
Multipath resistance	Multipath signal rejection to ease critical system use. (Multipath signals are processed in the same way as interference.)
Ranging	Location of units in a network with concurrent identification. (Auto-correlation of codes provides high resolution signal.)
Privacy	Protection of privileged communications from eavesdroppers. (Code signal structure requires code analysis to detect information.)

All of the above capabilities listed are dependent upon a coded modulation method for implementation. There are three basic signal formats that may be considered for use:

1. Direct sequence modulation - wherein a carrier is directly modulated by a code (usually some form of angle modulation with the code bit rate much higher than the information signal bandwidth).

2. Frequency hopping - in which a code is used to cause a carrier to jump from one frequency to another, generally pseudorandomly.

3. Chirp - under code control only in those cases in which the code determines sweep direction. Chirp or pulsed FM modulation occurs when a carrier is swept over a wide band during a given pulse interval.

The codes employed are the instrument through which the advantages of spread-spectrum communication are realized, not only providing transmit signal bandspreading, but acting as a key to despreading

in the receiver. A conventional receiver, not having the proper key, can neither receive a spread-spectrum signal, nor is a conventional receiver greatly affected by it.

For time division multiplexing of signals, the spread-spectrum waveforms provide accurate timing, since a synchronized transmitter and receiver track to within a fraction of a code "chip". (The term "chip" here refers to the frequency hopping rate.) Since typical code chip rates in these systems are in the range of 1 Mbps to 50 Mbps for direct sequence and 1 Kbps to 100 Kbps for frequency hopping, system timing is never worse than about 1 msec, and is often as good as few nanoseconds. This means that (except for allowances that must be made up for differences in signal propagation time) a whole network of transmitting and receiving stations can take turns using a single frequency assignment, by synchronously switching from receive to transmit or vice versa at the same time.

Code divison multiplexing of signals depends upon timing also, but in a different way. Either direct sequence or frequence hopping signals depend upon the codes used in generating their modulation to provide a low degree of correlation between the instantaneous frequency spectra of the signals transmitted. The prime difference in the time division and code division approaches is that time division implies coordination of time slots, while code division does not.

The frequency hopping format is such that a given transmitter, in jumping from one frequency to another, dwells at each frequency for only a small portion of the time available, and a conventional receiver sharing the same band hears only an occasional short burst of what is apparently (to it) noise. This is true, and must be realized, even though the transmitter may employ only very little power. By contrast, a direct sequence system transmits continuous noise-like interference to all receivers in the same band, although of a very low level. For this reason, frequency hopping is often preferable when there are "near-far" locations to be considered.

For comparison, in a frequency band having a direct-sequence spread-spectrum signal centered around an equal level narrow band signal, both receivers could operate with the narrow band receiver seeing a carrier-to-noise ratio of 23 dB and the direct sequence receiver rejecting the narrow band signal as interference.

In summary, the common channel usage capability of these systems is far from their only attraction. All of the capabilities listed above are readily available--privacy, discrete addressing, multipath rejection and multiple users on the same channel are all attractive features. Some current applications of spread-spectrum technology are described below.

The Global Positioning Satellite makes use of the spread-spectrum concept. In its operational environment, 24 satellites are to be orbited at about 1,000 nm to provide earth coverage for navigation and weapon delivery. At least six satellites are to be in view from any location. The satellites will broadcast their identity, position and also highly accurate time. User equipment would select four of the most appropriate satellites and solve four equations in four unknowns to display to the user his position in the dimensions and time. For the purpose of enroute navigation, non-military marine or aeronautical users could determine their position on a world-wide grid accurate to about 100-200 meters. Through the use of secret coding, military and NASA systems will have access to positioning at a much higher level of accuracy. For example, the space shuttle will use spread-spectrum systems for communication and navigation. Its position will be known to within 30 feet, its velocity to within 0.02 ft/sec and time to within a fraction of a microsecond (25).

Another spread-spectrum system is the Joint Tactical Information Distribution System (JTIDS) which will operate in the 960-1215 MHz band, which is allocated worldwide for the Aeronautical Radionavigation Service on a primary basis. The JTIDS system will operate in this band, except that it will hop over the 1030-1090 MHz region. The JTIDS concept provides for the integration of Air Traffic Control Radar Beacon System (ATCRBS), IFF, Discrete Address Beacon System (DABS) and Global Positioning System (GPS) capabilities.

The 960-1215 MHz band is divided into two subbands in the U.S.: 960-978 MHz is allocated for military use and 978-1215 MHz is allocated for Government and non-Government use to support the National Air Space (NAS) Air Traffic Control system. The 960-1215 MHz band is presently used by TACAN/DME, ATCRBS, and IFF beacon functions. The TACAN system provides an aircraft pilot with slant range and bearing to selected ground beacons, whereas DME provides only slant range. In addition, some of the General Aviation DME sets calculate the radial velocity based upon the rate of change in range.

The ATCRBS and IFF systems provide surveillance as well as identification information at 1030 and 1090 MHz. The beacon interrogator transmits on 1030 MHz and the avionics transponder replies on 1090 MHz. The JTIDS signal does not transmit on channels close to and including 1030 and 1090 MHz so that the potential for interference at these frequencies is minimized. The JTIDS system employs a time division multiple-access technique.

The OTP (NTIA) analysis (21) based on flight tests and field measurements indicated the JTIDS system could be co-located in this band without interfering with other systems if the JTIDS ground terminals are properly sited. As a consequence, the FCC has approved JTIDS for operating in this band which is shared by both Government and non-Government users.

r area, in 1976 the FCC initiated studies looking into ty of using spread-spectrum techniques to possibly heavily congested services, such as land mobile. s (22, 23) indicate that conventional FM is more cient than spread-spectrum modulation for a land mobile sting of many independent networks; however, it was t one or more spread-spectrum systems could be overlaid ng FM band without causing unacceptable interference to service and that spread-spectrum system performance in died is markedly better than non-spread-spectrum ultipath environments. It is clear that much more ork and field tests will be required in order to make rmination.

example of spread-spectrum applications is the concurrent of data in the same band as used by standard color TV increase the utilization of existing TV channels and to sed systems in which graphical and alphanumeric data mitted. Experiments show that data can be inserted sequence spread-spectrum modulation at a level that is e to the viewers, and yet may be recovered with an f less than 2 x 10^{-5} at data rates up to 100 Kbs (26).

5.4 Wire Transmission Systems

Wire is the oldest medium for electrical telecommunications and remains the backbone of the telecommunications network, particulary for local loops. Cables containing 17 or more twisted wire pairs are the dominant means of extending telephone service to the individual subscriber. The average length of these local loops is about two miles. The loop cost varies linearly with distance, however, and for loop lengths greater than about five miles other media become more economically attractive. Metallic coaxial cables offer bandwidths of several megahertz, allowing the multiplexing of many voice channels over a single cable in either frequency or time. Digital multiplexing is currently more attractive than analog for short distances (less than about 300 miles), where the cost of terminal equipment is greater than that of transmission equipment. For longer distances, analog transmission is less expensive. The crossover points are expected to shift somewhat during the 1980s so that the range best served by digital transmission will be extended in both directions (24).

There are many disadvantages of wire transmission systems, not the least of which is the limited bandwidth available for information transfer. The low information capacity of twisted pairs is a major obstacle in providing many of the broadband features mentioned in section 3 to the local subscriber. While electronic mail, funds

transfer, and perhaps news can be transferred over existing wire pairs, video information can not, at least not in real time. Slow-scan techniques can permit limited picture transmission which is probably adequate for graphics but for real-time video and high-speed data, a broader band local loop is required. Local cable systems are a possibility; however, the cable television industry has not yet expanded at hoped-for rates. In 1975 (27), 90 percent of all installations were rural (serving areas remote from broad-cast antennas) and fewer than one percent of the existing and planned systems provided for active two-way services.

The Bell System Picturephone service now uses a 5 x 5.5 inch display with 30 frames per second and 251 visible horizontal lines (of 211 elements each) per frame. The channel bandwidth is about 1 MHz. Transmission is over specially conditioned twisted pairs for short distances or 6.3 Mb/s digital links for longer loops. Even with this reduced format (compared with commercial TV), it has been estimated that a video telephone system with two million units installed in 1984 would require an investiment of roughly $20 billion. For a cable television-based two-way system, the investment would likely be less but still near $10 billion (28). The cable-TV concept usually considered includes video <u>to</u> the subscriber but only audio or low-speed data <u>from</u> the subscriber and is hence more limited. For reference, the Boeing 747 development costs have

been estimated at $750 million. By 1980, the pollution abatement program will have consumed an investment of $93 billion. Certainly, new communication functions will be extremely expensive on a wide scale and will have to compete with a variety of other programs for available capital.

A number of alternatives to the use of wire pairs in local subscriber loops are being investigated and in some cases they have already been implemented. In some metropolitan areas, cable duct space under the streets and in large buildings is becoming a premium, forcing the consideration of digital carrier technology instead of single wire pairs. An alternative solution is AT and T's recently deployed Loop Switching System which triples the number of customers served by existing voice-frequency cables. It is a multiplexing system based on the concept that only a few customers make calls simultaneously. Consequently, much of the time loops are idle and can be shared (29). Higher density local traffic may require the consideration of 1.544 M bit/sec digital carrier over coaxial cable, fiber optics, short microwave hops or even free-space laser links.

Telephone service in rural areas has traditionally been costly because of long distances between customers and the central office, requiring the use of heavy gauge cable to provide signaling and transmission. However, a subscriber loop carrier system using

digital technology is now in operation in several areas and is considerably more cost effective for providing such service than the older analog party-line technology. In this system, up to 40 simultaneous phone conversations can be handled on two pairs of wire and the subscribers can be located as far as 50 miles from the central office (30).

The advantages of wire pairs for local loop transmission have been their ability to provide voice-quality transmission bandwidth, signaling, and power to the subscriber's telephone at relatively low cost. The cost advantage of wire relative to fiberoptics is rapidly decreasing, however, due to the increasing cost of copper and the decreasing cost of the fibers. The tremendous bandwidth advantage of fibers, pressure for new features, and changing economic realities will likely result in the increasing use of optical fibers for new equipment installations during the 1980s. This trend will begin most strongly in PBX's and spread to local loops. The use of fiberoptic link to provide power and signaling for a solid-state telephone has already been demonstrated. Cable television is now switching from metallic coaxial cables to fiberoptics and this can be expected to continue. Whether this approach will pervade local loops as a supplement to the telephone wire-pair paths or whether the telephone network will itself assume the lead in converting the local plant to a broadband link is unclear. A significant decrease in fiberoptic cable, coupled with substantial increases in the cost

of copper, could provide motivation not only for fiberoptic use in new systems but for the replacement (and recovery) of existing copper loops as well. In any case, wire is likely to be replaced by fiberoptics for new business (and ATC) applications during the next decade and broadband cables should pervade the home subscriber loops during the 1990-2010 period.

The use of noncoherent infrared links has been recently announced as a replacement for premises wiring used for connecting voice/data terminal equipment to the telephone network or to on-site computers. Datapoint Corporation's LightLink™ infrared system is capable of free-space operation up to two miles at data rates up to 4 Mbs. IBM's Zurich Laboratory recently developed a 64 Kbs infrared data link for connecting data terminals in a working area to a central infrared LED station mounted on the ceiling of the room; and Siemens announced infrared cordless links for use with telephones and an infrared data link for process control operations in industrial plants. (These systems are free from electromagnetic interference generated by motors, arc welders and other RF sources). (31, 32).

5.5 Fiberoptic Transmission Systems

Fiberoptics is probably the most active and exciting area in telecommunications and will likely do more to change telecommunications over the next forty years than any other single development. As a replacement for conventional coaxial cable and wire pairs,

fiberoptic systems offer many advantages. Optical fibers offer between two and four orders of magnitude greater transmission bandwidth than do metallic cables, and this is probably their most attractive feature. In addition, however, they offer smaller size, less weight, less cross-talk, immunity to inductive pickup, greater strength, lower attenuation, and potentially lower cost. Such fibers have the potential of being applied wherever metallic cable is now used--in undersea cable, long-haul buried cable, underground short-haul cable, local loops and equipment wiring. For example, the 450 pounds of copper wire now used in fighter aircraft could be replaced with only 50 pounds of fiber cable, and in the Navy A-7 aircraft, 13 fiberoptic cables have already supplanted 115 wire channels representing 302 separate conductors (33).

Light-wave communication was first demonstrated by Alexander Graham Bell's "photophone" in 1880, but the practical application of the light-wave communication concept was not possible until about ten years ago for lack of a suitable light source and low-loss transmission medium. Kao and Hockman (34) pointed out the feasibility of using a dielectric fiber as a waveguide for data transmission at optical frequencies; however, practical telecommunications was not possible until Corning developed a glass fiber with attenuation of less than 20 db/km in 1970. In the decade which followed, these fibers were to see their loss drop by more

than a factor of five, with attenuation below 4 db/km in 1979 for some fibers. It is expected that worldwide sales of fiberoptic assemblies for communications will increase at an average annual rate of 50 percent from 1985-1990. The total sales, $39 million in 1978, should reach $659 million by 1985 and $1770 million by 1990 (35).

Nearly all of the present fiberoptic systems operate at a wavelength of about 850 mm (0.85 μ m) at relatively low data rates usually less than 100 Mb/s. During the next five years, substantially longer wavelengths and higher data rates should appear. Most present North American systems link central offices at 44.7 Mb/s; however, two Canadian installations are used to implement two-way subscriber loops. One such loop connects 45 customers, providing telephone service and, at a test site, interactive cable TV (CATV) (36). The widespread use of fiberoptics in local loops is technically possible now; however, it is unlikely to occur very soon, both due to the high cost mentioned earlier and to the territorial conflicts between telephone and CATV companies which will have to be settled. Another pioneering experiment in Canada is a 30-mile long T4 (274 Mb/s) fiberoptic trunking system sponsored by the Alberta Government Telephones and expected to be in operation in late 1979. The system will use a 2-mile nominal

repeater spacing and cable with an 8 db/km loss (including any splices). The bit error rate is 8.5×10^{-11}; jitter, 0.03 ns(rms); and system availability, 99.999%. The dominant system reliability concern is the injection laser diode that must have a minimum mean-time-between-failures (MTBF) rating of 60,000 hr to meet the system reliability requirements. This system will have automatic channel protection switching capability and an automatic fault isolation circuit (36, 37).

In Japan, systems for telephone voice, high-speed data, and television (CATV, TV, high resolution TV) are rapidly being developed. One system, under test in Tokyo, features cables having 2.8 db/km loss at 0.85 μm and an average bandwidth of 800 MHz/km. Transmission rates of 32 and 100 Mb/s have been used. Laboratory tests using 1.3 μm transmission on single-mode fibers at 800 Mb/s are being conducted (38). Vigorous activity is also occurring in Europe, where data rates as high as 1120 Mb/s (1.12 Gb/s) are under exploratory development (39).

Within the past five years a number of optical fiber transmission systems have been installed by the Bell System and several independent telelphone companies, mainly as interoffice trunks. As the result of its sucessful experiments in Chicago and Atlanta, Western Electric is now in the process of developing the FT3 44.7 Mb/s system for use by the Bell System operating companies, which will have from 12 to 144 fibers in one cable. The largest configuration will have

44,352 message channel capacity, with 7 km repeater spacing. The FT3 system will have automatic performance monitoring and protection switching features and its main use will be for metropolitan trunking (40).

Investigations are underway exploring the use of fiberoptics for submarine cables, backplane wiring on large switches, and entrance links to satellite antenna facilities. A possibility also exists for expanding channel capacity by wavelength multiplexing in the 800-1200 nanometer optical band. Beyond 1983, fiberoptic cables are expected to control most of the market for inter-exchange cable in the Bell System and in the independent telephone companies (41).

The present state-of-the-art for fiberoptic systems can be considered first generation systems. It is not difficult to envisage second generation systems in which signal processing, including switching and possibly amplification at the repeaters and terminals, will be done at optical frequencies. This development may prove an efficient method for greatly increasing system bandwidth and avoiding the problems of electronic circuit response time. Integrated optical components will enable the capabilities of the fiber as a transmission medium to be more fully realized. In fact, the repeater design eventually may be simpler since conversion to electronics can be eliminated, power consumption may be reduced and system reliability improved (42).

The full capabilities of fiberoptic systems will not be realized until the advent of the third generation systems in which video or speech will be directly converted to optical signals by means of video or acoustic optical transducers. Such a fundamental attack on the subscriber equipment and its compatibility with the transmission medium may well result in a fundamental cost reduction by simplifying the presently required signal conversion process (42). While integrated optical devices are under development today, second generation systems are not likely before the 1990s, and third generation systems should occupy us well into the next century. There is little doubt, however, that fiberoptic technology will provide a major impetus to develop integrated optical devices for communication in the twenty-first century.

In the remainder of this section, the major components of a fiberoptic transmission system--the transmitter, fiber, and receiver--will be discussed and progress will be forecast to the year 2000.

5.5.1 Optical Fibers

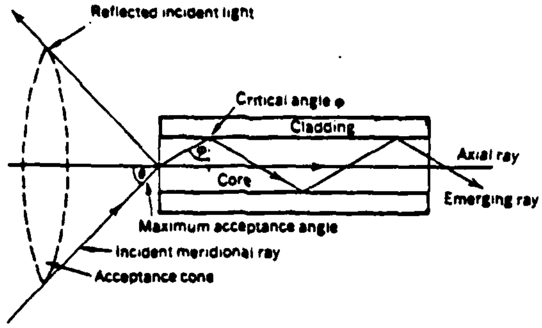
Fiberoptic transmission systems consist of three distinct parts: a tansmitter, which converts the electrical signals into an optical signal in the red or near infrared; the optical fiber itself, which provides a path for the light; and a receiver, which

converts the optical signal back into an electrical one. Any one of these three components can limit the cost or bandwidth of the system, and the success of fiberoptics is the result of progress in all three areas. This section considers the optical fibers themselves, while subsequent sections discuss the transmitter and receiver functions.

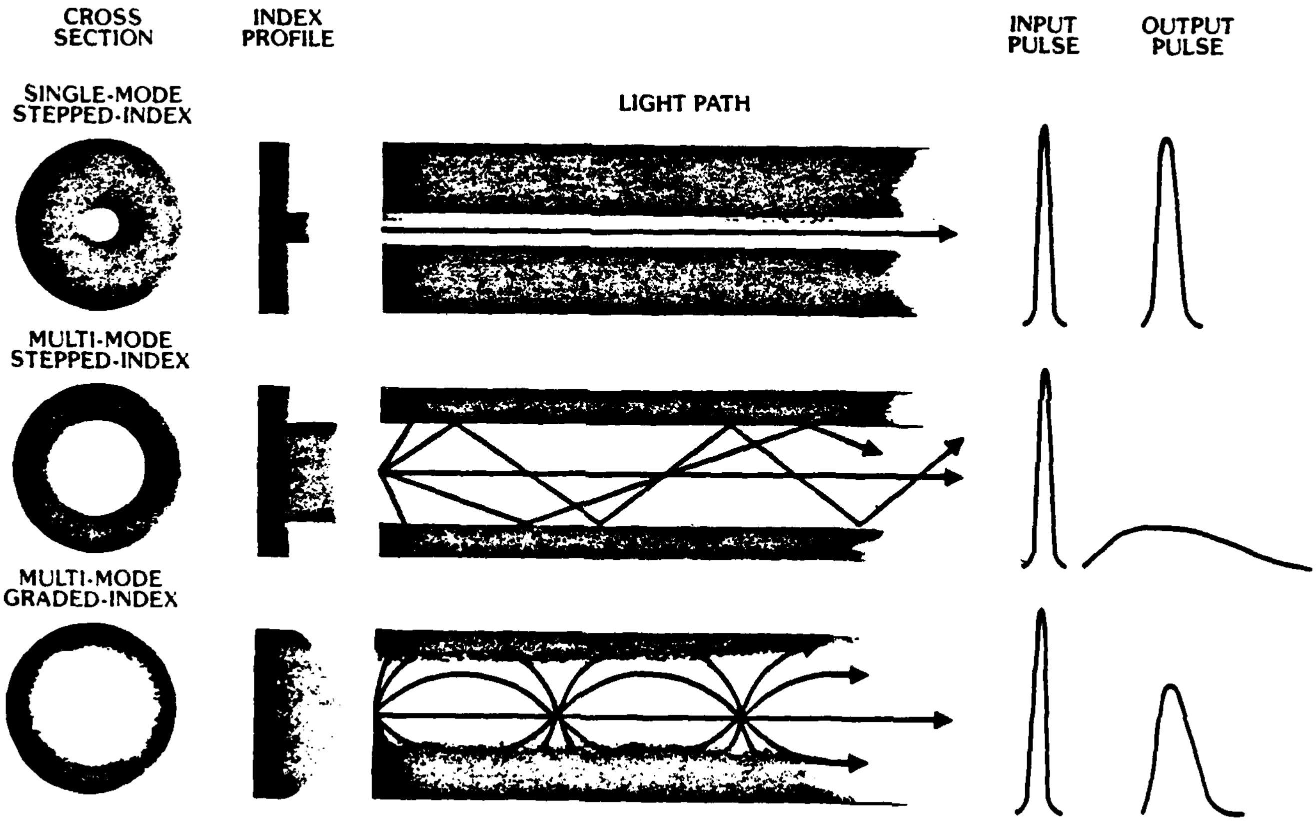
Optical fibers consist of thin strands of glass (silica) or plastic having core diameters typically less than 100 μ m (about the size of a human hair). The function of the fiber is to transmit light with a minimum of loss or dispersion. Light is contained within the fiber via the principle of total internal reflection, which states that a ray striking the boundary between two materials having different indexes of refraction (m) will be reflected completely back into the first medium, provided its angle of incidence is less than a critical angle ϕ, given by

$$\phi = \sin^{-1}(m_2/m_1)$$

where m_1 and m_2 are indexes of refraction. This situation is shown in Figure 5.6 (43), where m_1 represents the fiber core and m_2 represents the cladding. The critical angle also determines the maximum acceptance angle for the fiber, the sine of which is the numerical aperture (NA). This parameter (NA) can be used as a measure of the light-gathering ability of the fiber, and present NA values range between 0.1 and 0.5. Light arriving within the acceptance cone will travel down the fiber, being totally reflected by the walls and exiting only at the far (receiver) end.



(a)



(b)

Figure 5-6: Optical Fiber Types and Characteristics
(a) Definition of Acceptance Angle and Critical Angle
(b) Three Fiber Types with Profiles and Propagation Characteristics

(43, 49)

ACUMENICS

Optical fibers are classified by their index of refraction profile and modal properties. The three prominent types are the step-index multimode fiber, the graded-index multimode fiber, and the single-mode fiber. Figure 5.6 also illustrates these differences in the refractive index profiles.

The step-index fiber consists of a core (of doped silica glass) surrounded by a second material that provides the reflective boundary. If the second material is a layer of glass or other solid material it is called the cladding; unclad fibers have air as the second medium. The core is designed with an index of refraction greater than that of the cladding, which in accordance with the above equation (Snell's law), defines an appropriate critical angle. In step-index multimode fibers, each mode has its own characteristic velocity, and propagation delay differences give rise to signal dispersion (pulse spreading) which limits transmission rates to about 50 Mb/km. Core diameters are also relatively large and attenuation is high, while costs are low.

The profile for the graded-index fiber differs markedly from that of the step-index fiber. The index of refraction for a graded-index fiber changes gradually (usually in a radially symmetric, parabolic shape) instead of abruptly. The result of this index grading is to control the propagation so as to equalize the path

lengths of the various fiber modes and minimize dispersion. This is accomplished by a continual refocussing of the light rays as they travel down the fiber; hence, the graded-index multimode fiber acts as a lensing medium (44). This lensing action increases the bandwidth (information capacity) of the graded-index fiber between one and two orders of magnitude relative to the step-index fiber.

Single-mode propagation is realized by designing core diameters to within a few wavelengths of the light signal, and by having small refractive index differences between the core and the cladding. Although structurally similar to the step-index fiber, the single-mode fiber exhibits superior performance, even when compared to graded-index fibers. The single-mode fiber offers very low loss and extremely large bandwidths (in excess of 1 Gb/km). However, single-mode fibers are difficult to work with. Fiber shape, refractive index, and operating wavelength must all be carefully controlled.

An attribute common to all data links is attenuation. Repeaters can, of course, be used to upgrade signal quality; however, they add to total system cost and complexity. The principle causes of attenuation in optical fibers arise from <u>intrinsic absorption</u>, <u>vibrational (infrared) absorption</u>, <u>impurity absorption</u>, and <u>Rayleigh scattering</u>.

Intrinsic absorption involves the absorption of energy from the propagating light wave by the atoms that constitute the fiber. This absorption is most profound at ultraviolet wavelengths (45). Vibrational absorption results from the energy loss that occurs when the incident light wave excites the fundamental vibrational mode of the core lattice structure. Impurity absorption occurs when impurity ions, usually OH radicals from water, absorb energy from the incident light. Other significant impurity ions include the transition metals, cobalt, chromium, and copper. To insure that losses due to absorption by impurity ions are less than 1 db/km, all impurities must be reduced to a concentration in the parts per billion range (43, 44).

Rayleigh scattering occurs when light particles collide with imperfections (or irregularities) in the core's lattice structure. The imperfections arise from phase and compositional fluctuations frozen in during glass melt/cooling; imperfections also arise from frozen-in thermal fluctuations. As in all Rayleigh scattering, this loss decreases rapidly with increasing wavelength. Rayleigh scattering repesents the intrinsic lower loss limit, which is about 0.2 db/km (44).

Theoretically, there are two wavelengths where the attenuation of silica glass fibers reaches very low values. The first is between 1.2 and 1.3 microns, and is due to the decreased importance

of ultraviolet absorption loss. The second loss minimum occurs between 1.5 and 1.6 microns, and is due to the reduced importance of infrared absorption loss. These low loss regions are attractive targets for the transmission of data over long distances.

Figure 5.7 (a and b) compares the attenuation-frequency characteristic of optical fibers with several types of metallic coaxial cable. The fiberoptic response is relatively flat compared with the roughly square-root dependence of coaxial attenuation on frequency (46). Figure 5.7 (c and d) compares the attenuation-wavelength and attenuation-frequency (bandwidth) characteristics of step-index and graded-index fibers (47).

There are three main types of signal dispersion (pulse spreading) in optical fibers: waveguide, material, and intermodal. Waveguide dispersion results from the close relationship between the mode and wavelength of the propagating signal. Waveguide dispersion alone is insignificant because it affects only the highest modes. Its importance surfaces when its relationship with other types of dispersion is considered.

Material dispersion arises from the dependence of the group velocities of the modes in a propagating wave on wavelength. Single-mode fibers are particularly sensitive to material dispersion.

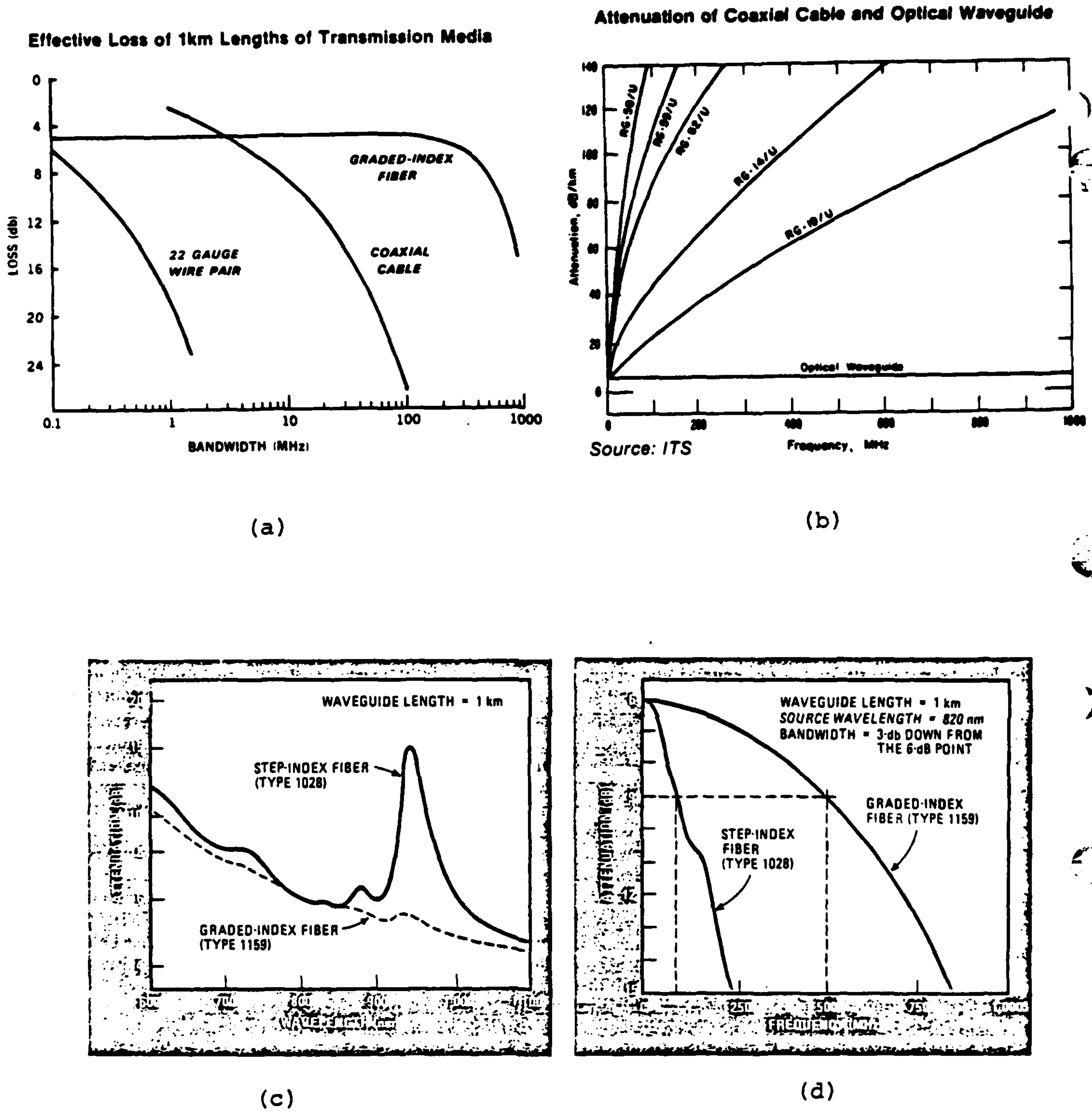


Figure 5.7: Attenuation Characteristics of Optical Fibers (47, 49)

Above 1.27 microns, material dispersion (for silica) changes sign relative to its value below 1.27 microns. This sign change is important because at some value near 1.27 microns, material and waveguide dispersions completely cancel each other. Here information capacity can increase dramatically, because signal pulses may be spaced close together without fear of the interference caused by pulse spreading. Theoretically, single-mode fibers can achieve bandwidths in excess of 100 GHz/km (44). Since fiber attenuation is also a minimum near 1.3 microns, this is a particularly attractive wavelength for data transmission.

Intermodal dispersion in multimode fibers arises from interference among the many modes present in these fibers. Single-mode fibers are much less sensitive, and modal effects may be disregarded. Step-index fibers are very susceptible to intermodal dispersion and their information capacity is significantly lower than that of graded-index fibers as a result. Intermodal dispersion is highly dependent upon refractive index; hence, careful control of fabrication techniques must be maintained to insure proper operation.

Mode mixing is an effect which reduces dispersion in multimode fibers, especially for longer fibers. It involves the scattering of light at structural imperfections, which reduces dispersion by averaging all mode velocities at the output. Mode mixing is also a source of attenuation due to the energy lost in the scattering process (45).

Further comments on dispersion will be given in the next section on sources, since the spectral purity of the source, as well as the numeric aperture of the fiber, is important in determining the dispersion of the system.

In addition to attenuation and dispersion, other important parameters associated with fibers are strength, the ability to be spliced with low loss, and of course, cost. Fibers have a tensile strength which is theoretically higher than metal. They bend easily and are generally easier to install in buried conduit than are coaxial cables, partly due to their lighter weight. Fibers are formed in cables such as that shown in Figure 5.8 (43), which contains 144 separate fibers.

Connectors and splices are particularly important in fiberoptic systems, since a slight misalignment can cause high loss, and several splices/connectors are required in a typical transmission path. The ability to be accurately spliced in the field is particularly important. To maintain loss/splice at about 0.2 dB, the fiber must be aligned to within 0.2 core radius. For multimode fibers this corresponds to a distance of more than 10 μ m which can be approached at low cost. For single mode fibers, 0.2 core radius is less than 1 μ m, creating a connector/splice problem that may be expensive to solve (43).

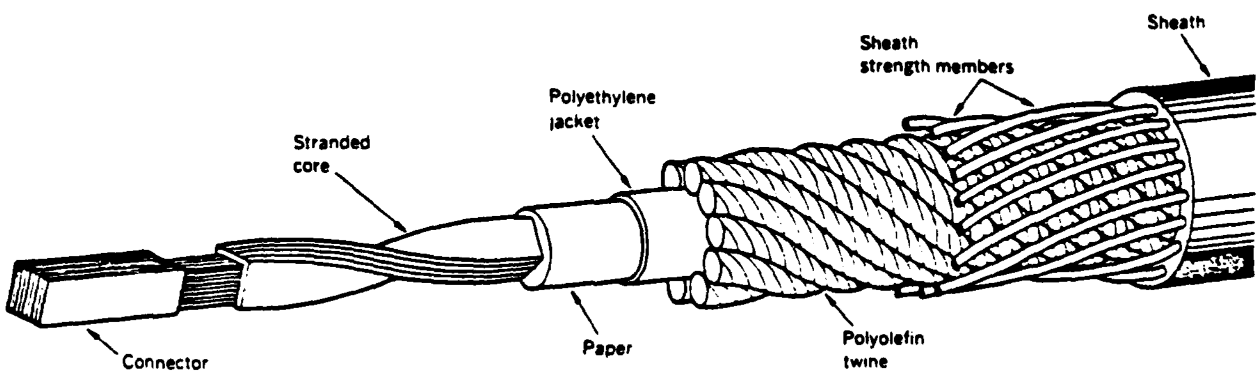


Figure 5.8: Fiber-Optic Cable

This Fiberoptic Cable Design consists of 12 stacked ribbons, each containing 12 glass fibers. The sheath strength members are steel wires. Cables of this type were used in the Bell System's Atlanta trial. (43)

The cost of fiberoptic cable on a per-channel (per fiber) basis has been dropping rapidly and should continue to do so during the 1980s. Cost will depend on fiber quality (attenuation, bandwidth, dispersion), on the number of fibers/cable, and on the quantity ordered. For single-fiber cable having loss in the 3-6 db/km range and a bandwidth exceeding 500 MHz/km (a graded-index fiber capable of TV transmission), the cost was about $2.50 per meter in 1976 compared to between $0.50 per meter and $5 per meter for various coaxial cables (33, 47, 48). By 1979, the cost of the fiberoptic cable had dropped to less than $1 per meter (49). Figure 5.9 shows the forecast cost of such cable between 1980 and 2000. During this time the annual volume of cable produced should increase by several orders of magnitude while the cost per meter should drop by a factor of at least 20. Compared with present coaxial cable on a per-channel basis (number of voice circuits), the cost per voice channel should drop by more than a thousand over the next twenty years. Single-mode fibers should be well developed by 1990, allowing data rates in the 2-4 Gb/s range with loss of less than 1 db/km at 1.3 μ m.

5.5.2 Transmitters (Sources in Fiberoptic Systems)

The optical source is a critical component in fiberoptic transmission systems. It must generate a high output power of high spectral purity at a suitable wavelength at low cost and be able to switch rapidly to permit high bit rates to be launched into the fiber. Two types of sources are now used: light-emitting diodes (LED's) and laser diodes.

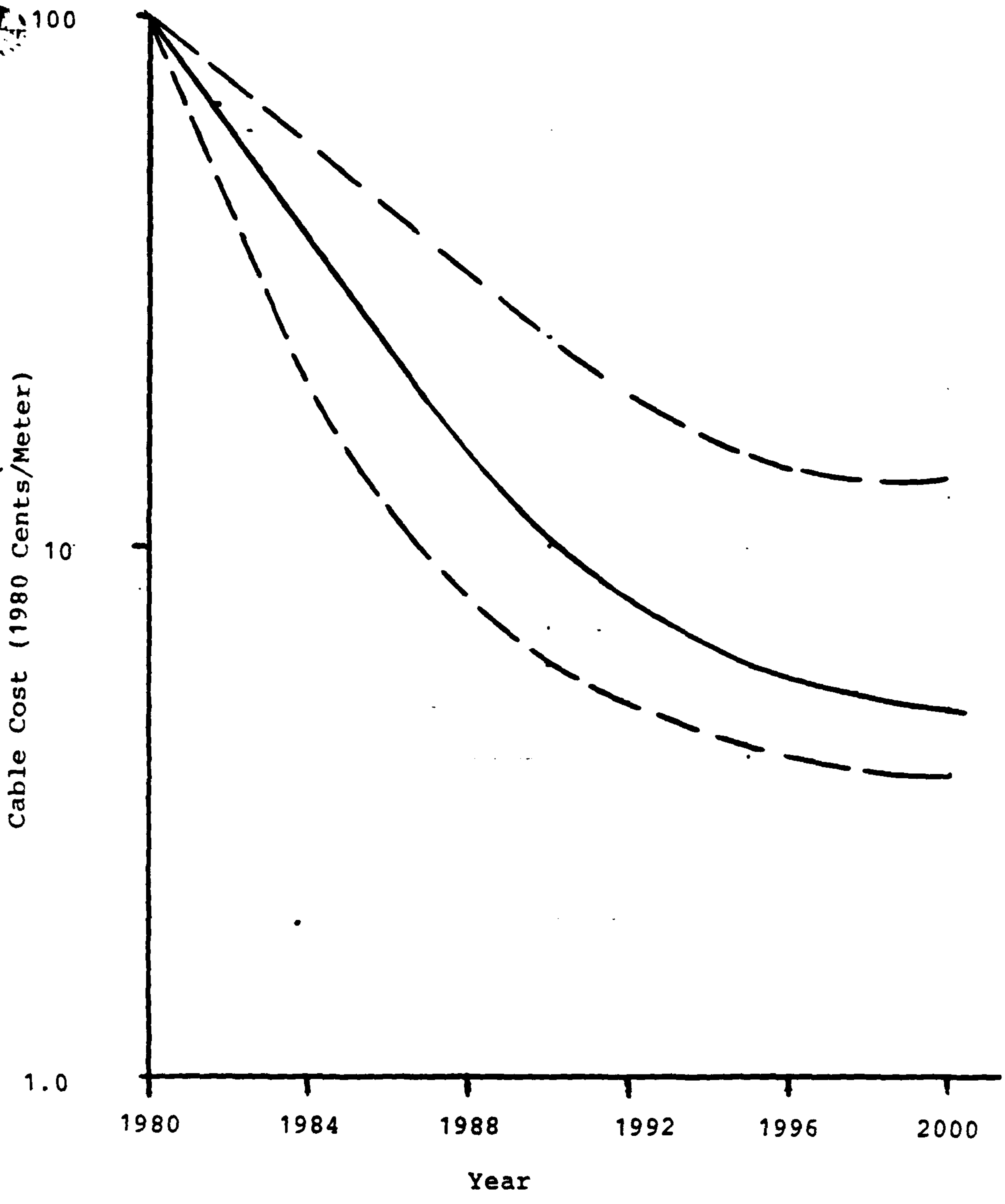
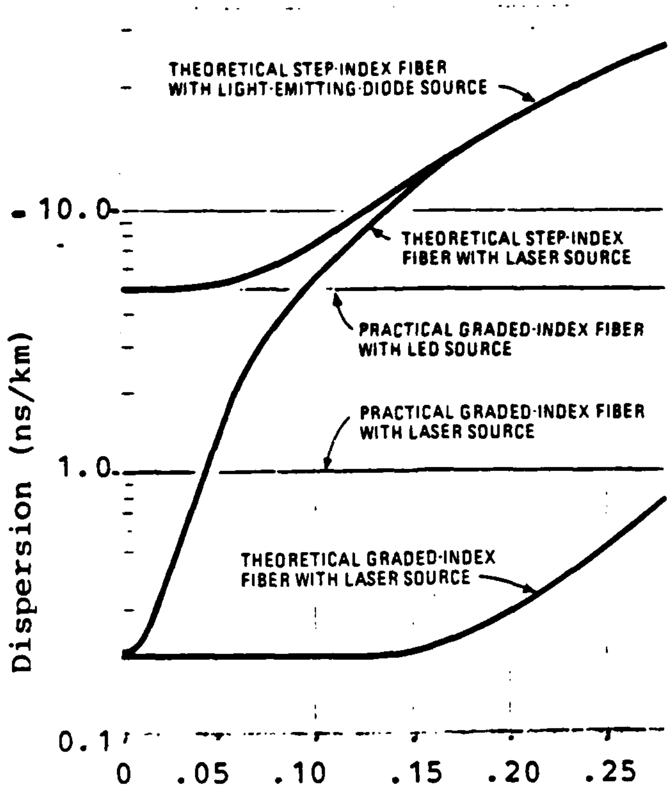


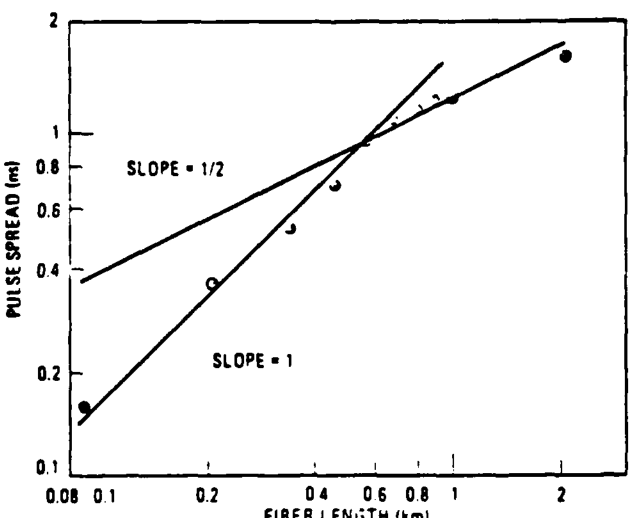
Figure 5.9: Cost Forecast for Single-Fiber Optical Cable Having Loss in the 3-6 db/km Range and a Bandwidth Exceeding 500 MHz/km.

The source must not only generate a high output power, it must also couple that power efficiently into the fiber. If this were the only criteria (which it might be for a short loop or low data rate), fibers with a high numerical aperture would be favored. However, at higher apertures (above 0.2-0.25), the wider acceptance angle results in increased dispersion and increased pulse spreading, which is a limitation on loop length or bit rate or both. Figure 5.10 (a) shows dispersion as a function of numerical aperture for step and graded multimode fibers with LED and laser sources (48). Practical levels are also indicated. The differences noted between source types results from the spectral purity of these sources. As noted in Figure 5.10 (b), dispersion is relatively large (1-4 ns/km) for LED's, which have, for example, a spectral width of about 0.04 μ m, but can be under 10Å (1 nm) for some lasers (46). Figure 5.10 (b) also shows the 3 db information rate limit imposed by the dispersion.

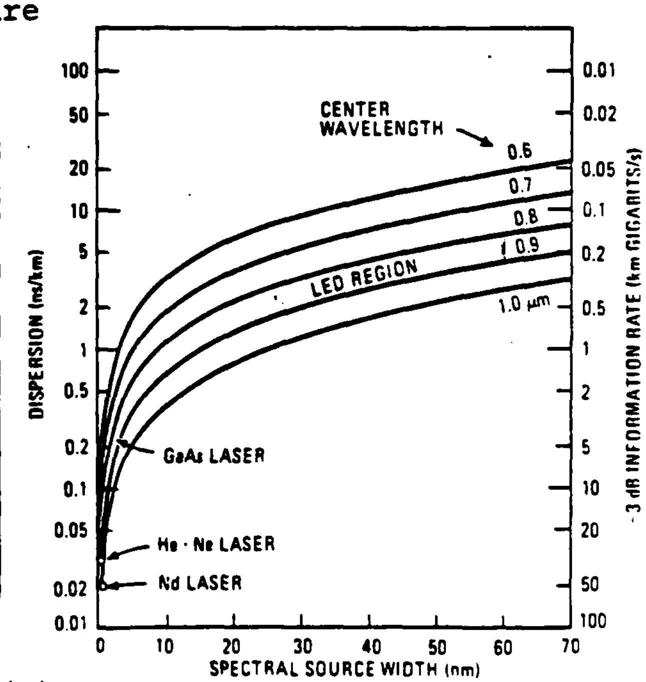
The most promising LED structures at present are double-heterojunction devices consisting of GaAs-GaAlAs-GaAs or InP-GaInAsP-InP sandwiched layers. Wavelengths, which have usually been in the 0.8-0.9 μ m range, now extend into the 1.05-1.3 μ m lower-fiber loss range (50) and are compatible with a dispersion limit above 1GHz-km. At 0.94 μ m, external efficiencies of 27 percent have been achieved with 54 mw of output power (51).



(a)



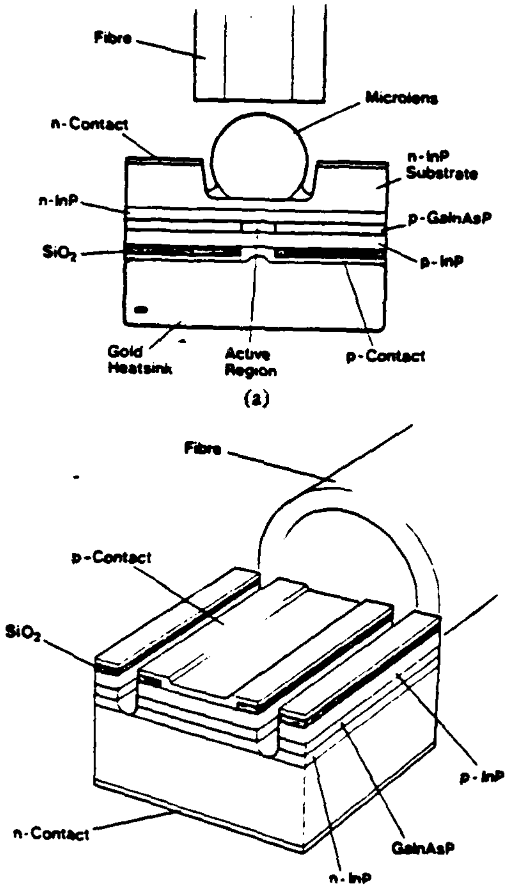
(b) Losses over long transmission distances can be minimized if a fiber's index of refraction is tapered properly from core to cladding. The loss is proportional to cable length over short distances to the square root of the length beyond 5 kilometers.



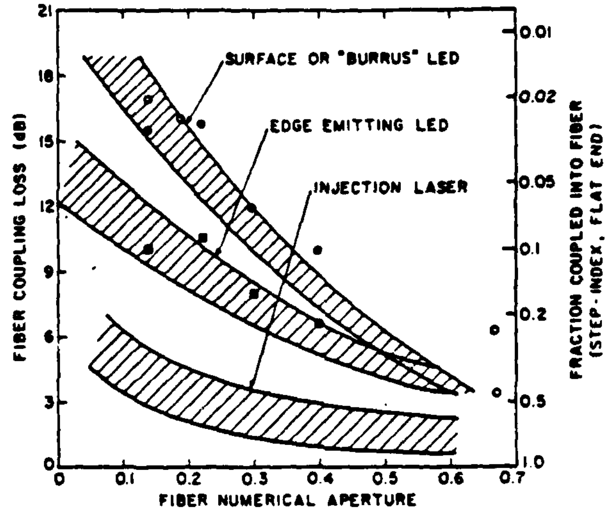
(c) Both geometric and material dispersion spread an optical pulse as it propagates down a fiber. Dispersion is a function of center operating wavelength and the source optical bandwidth

Figure 5.10: Optical Fiber Dispersion Characteristics (46,48)

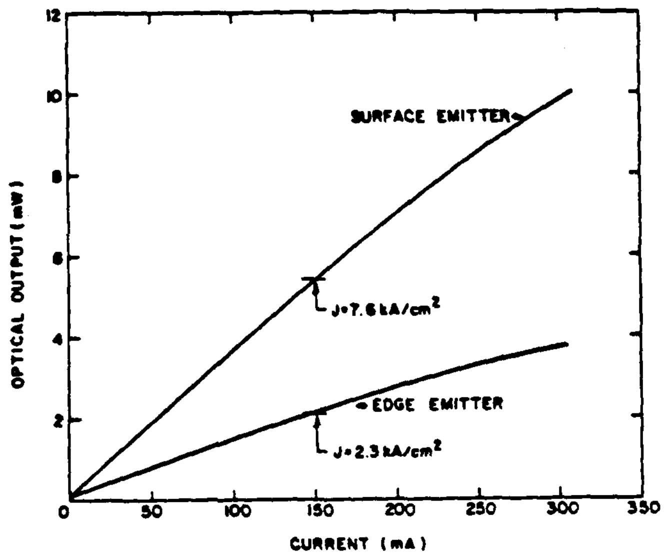
Figure 5.11 (52) compares the coupling loss of surface- and edge-emitting LED's with injection lasers. The differences are due to the directionality of these three types of sources. The optical output power and coupled power (at 150 mA drive) into a NA = 0.28 fiber are also shown. Edge emitters have larger bandwidths than surface emitters and can couple more power into low NA fibers. They are preferred for high data rates. Frequency response is related to material lifetime, which can be decreased at the expense of quantum efficiency. For data rates about 100 Mb/s, loss increases appreciably. LED's with output power in the range of a few milliwatts and bandwidths of 100 MHz or more are still relatively expensive ($200 in 1979) (49), but are expected to drop quickly, reaching $30 by 1985 and $5 in the year 2000. Higher efficiency, higher wavelengths, and narrower spectral widths, should also evolve, with price depending on performance. Extrapolated data from accelerated aging tests indicate LED's should provide more than 10^5 hours of operation (20 years = 1.75×10^5 hours) (53). LED's are suitable for analog intensity modulation as well as digital pulsing. They promise to be extremely useful as reliable low-cost sources for applications where the information rate-distance product is less than a few hundred Mb-km/s. While greater efficiency and longer wavelength are important goals, spectral purity is their prime limitation or long-distance high-speed communication.



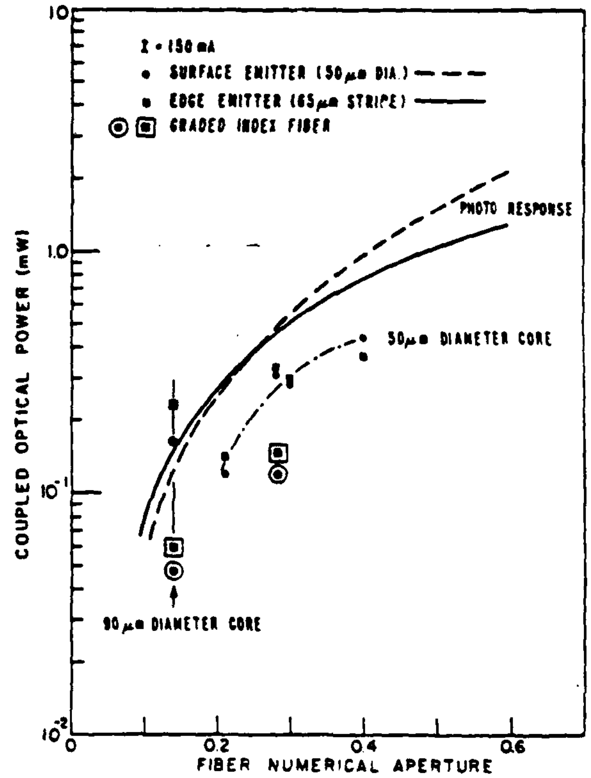
(a) Surface and Edge-Emitting LEDs



(b)
Relative Coupling Loss to Optical Fibers of Various Numerical Apertures



(c) Output Optical Power versus dc Drive Current for Surface and Edge Emitters



(d) Coupled Optical Power versus Numerical Aperture for the LEDs in (c).

Figure 5.11: A Comparison of LED Sources for Fiberoptic Transmission Systems (52)

Double heterostructure injection laser diodes have high spectral purity and narrower spatial output distributons than LED's. They produce output power in the range of tens of milliwatts with subnanosecond rise times and bandwidths about 1 GHz (49, 54). Their principal disadvantages are high cost ($500-$1000 in 1979) and lower reliability. If reliability problems can be solved, these structures should permit the full bandwidth of single-mode fibers to be utilized. Present data indicates lifetimes of more than 10^5 hours are achievable (43).

5.5.3 Receivers (Detectors) for Fiberoptic Systems

The receiver must detect the optical signal and convert it back to an electrical signal. Since repeaters add expense to the transmission path, sensitivity and low noise in the detector is extremely important, since it means wider repeater spacings. Hence, the figure of merit for a receiver is the minimum required input optical power needed to achieve a given level of performance (error rate or signal to noise level) at a given information rate. Silicon photo-diodes are used as detectors. Quantum efficiencies (ratio of primary photoelectrons generated to photons incident on the detector) of the order of 90 percent, and response times of 0.1 nsec are [illegible]able for wavelenghts in the 0.8-0.9 μm range (43). Unfor[illegible] silicon is increasingly transparent in the infrared, so [illegible] falls off quickly at longer wavelenghts.

Figure 5.12 (43) compares the sensitivity of PIN diodes and avalanche photodiodes (APDs), the two types used, as a function of bit rate. Avalanche photodiodes provide internal current gain (via avalanche multiplication) and greater sensitivity. Both devices, at present, are one to three orders of magnitude worse than the theoretical minimum. Both diode types are relatively inexpensive (less than $1). APDs require more sophisticated electronic amplification, however, including a high voltage supply (200 volts). Complete detectors now sell for $5-$50 for PINs and $50-$500 for APDs, depending on performance (49). The peripheral electronics should become more highly integrated and much lower in cost. A PIN detector with a bandwidth of 500 MHz which sells for $50 in 1980 should drop to $20 in 1985 and $7 in 2000.

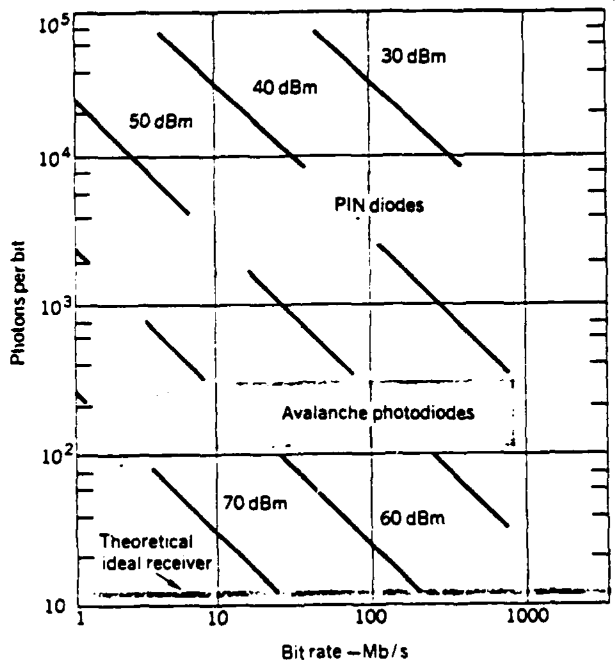


Figure 5.12: Sensitivity of Fiberoptic Receivers (43)

5.5.4 Concluding Comments on Fiberoptic Systems

Progress in fiberoptics is likely to continue at an accelerating rate. For long-haul trunks, single mode fibers should be capable of data rates exceeding 2 Gb/s by 1990; the cost per voice circuit is expected to drop from three to four orders of magnitude compared with present coaxial cables. The installation of an optical toll network will take considerable time, however, and will still be occurring in year 2000.

Figure 5.13 shows the forecast cost of a relatively modest 100 MHz fiberoptic link having a length of 10 km. The system is assumed constructed of 4 db/km graded-index multimode fiber with repeaters spaced at 2.5 km intervals. Relatively inexpensive LED and PIN sources and detectors are used. The total cost for repeaters (including connectors) never exceeds 15 percent of the cost of the link which is dominated by the cost of the cable. The link cost drops by nearly a factor of twenty between 1980 and 2000. It is noted that repeaterless 10 km links have already been used at 44.7 mb/s using laser sources and avalanche photodiode receivers (33). The forecast link, therefore, represents a relatively modest low-cost medium-traffic optical data path which should be widely used during the next twenty years.

5.6 Microwave and Satellite Transmission Systems

In parallel with fiberoptic developments, telecommunications will see the further development of microwave ground-to-ground and

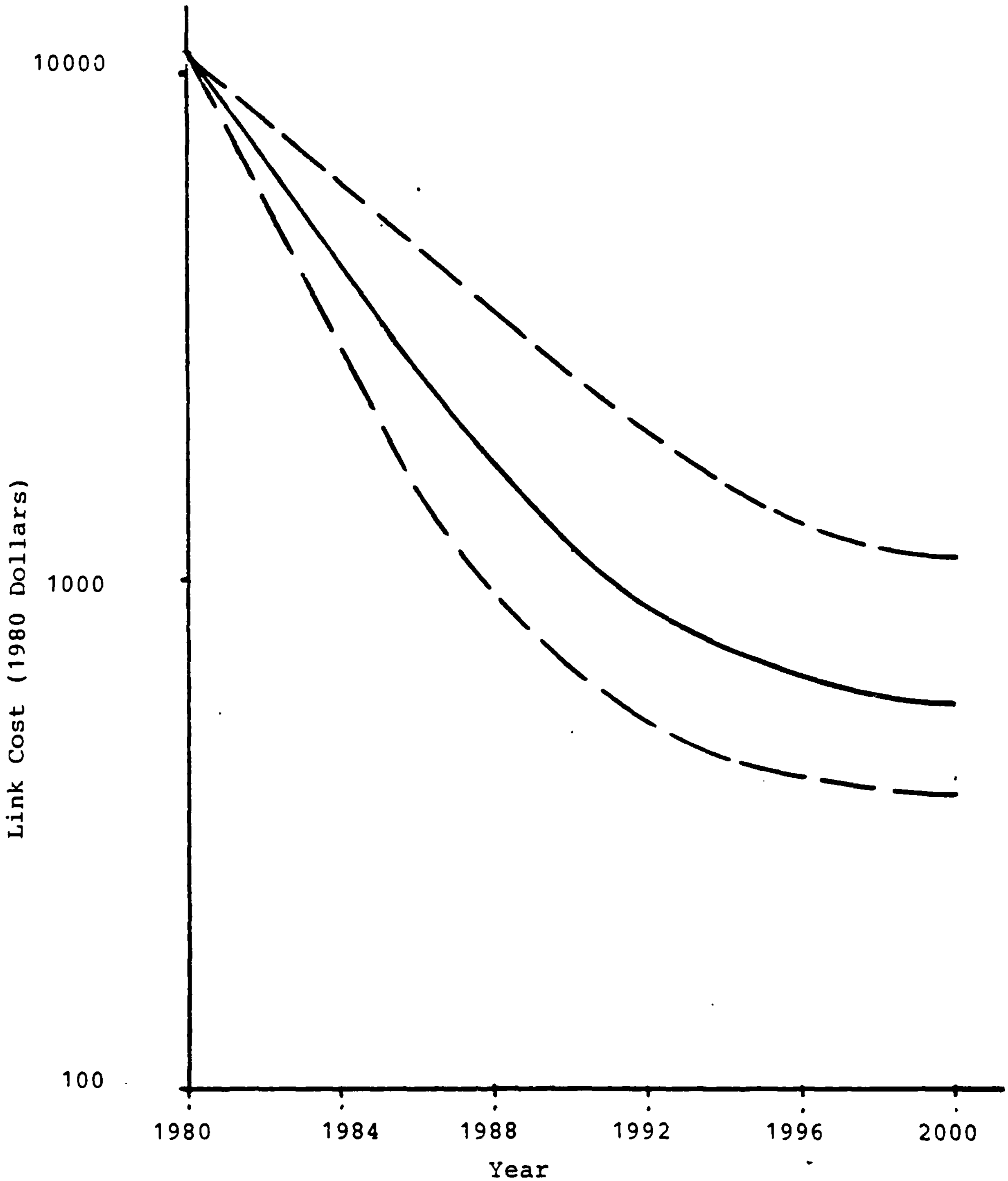


Fig. 5.13: Cost Forecast of a 10 km Fiberoptic Data Link. Graded-Index (4 db/km) Cable with LED and PIN Sources and Receivers.

ground-to-satellite transmission links. Fiberoptics could actually lessen the demand for microwave ground-to-ground channels; nevertheless, there will be increasing pressure for additional ground-satellite links to serve the needs of intercontinental traffic as well as mobile (e.g., airborne) applications. In developing nations, where extensive ground lines are not in place and where terrain may make their installation difficult, a greater emphasis will also be put on satellite communications.

This section will first briefly review microwave and satellite development and near-term expectations. This will be followed by a forecast of satellite communications for the 1980-2000 time period. Several further discussions of satellite communications can be found in the literature (56,57).

5.6.1 Microwave Transmission

In 1979, microwave radio-relay routes carried nearly 70 percent of the nationwide telephone carrier circuit mileage. In the Bell System, the TD-type analog system is the most prominent, representing an older technology utilizing the 4 and 6 GHz common carrier bands and the FDM-FM modulation scheme with a capacity of 1200 to 1500 circuits per radio channel. The TH-type systems operate in the 6 GHz band and have a capacity of 1800 voiceband circuits. The TH systems are used as an overbuild on existing TD routes, adding to the route capacity while utilizing existing structures and facilities (10).

A new ultralinear traveling wave tube has made possible the development of the type AR6A single-sideband (SSB) system, which has the capacity of over three times that of the TH system (6000 voice circuits vs. 1800 voice circuits). SSB modulated signals are more seriously affected by fading than are FM signals; however, the type AR6A system compensates for fading by use of special automatic again control circuits and, in the case of deep facing, a space diversity technique is used. (With space diversity, there are two receiving antennas, one on top of the other. When the signal from one antenna drops below a specific level, the receiver automatically switches to the other antenna, which is normally receiving a stronger signal.)

The incremental costs for adding new analog channels to existing routes making use of the AR6A SSB system are estimated to be about a dollar per circuit mile in contrast to an estimated \$5-\$10 circuit mile for a newly installed digital radio or coaxial system. A further reduction in circuit costs by a factor of two for this analog system is expected when the analog time assignment speech interpolation (TASI) system is introduced. (TASI systems are a type of multiplexing system which can increase the use of a given circuit by a factor of two.) As a consequence, long haul intercity voice traffic using analog carrier technology is expected to be cost effective for at least another decade (17).

Approximately 20 percent of the nationwide circuit network makes use of L-carrier coaxial analog transmission systems (chiefly, the L4 and L5 systems). These are SSB AM carrier systems having 32,400 and 108,000 channel capacities per route, respectively. The L4 system has a top frequency of about 18 MHz and the L5, about 65 MHz using 3/8-inch coaxial cables. Research appears to indicate that a system with a 150 MHz top-channel frequency is possible on 3/8-inch coaxial cables (10).

In 1974, only two types of digital transmission systems were available. Since then, six new digital systems have been introduced and two more are in the experimental stage. With the advent of digital switching systems, cost savings result as the number of analog-to-digital conversions are reduced. Time division multiplexed (TDM) carrier systems have the following advantages:

1. Use of low-cost solid-state circuits.
2. Elaborate lineup adjustments common to analog systems are not required.
3. A variety of services can share a line without the need to consider the requirements of the most sensitive service as is the case with some analog services.
4. Error rates are not a function of system length.

The T1 digital carrier was introduced in the early 1960's and was the initial short haul digital transmission system used on circuits up to 50 miles in length. (T1 accounts for approximately 8 percent of the total nationwide circuit mileage.) It carriers 24

voicegrade channels (1.544 Mb/s) over wire pairs with repeaters spaced about one mile apart and has a system bit error rate objective of 10^{-6}.

Additional T carrier configurations include the T1C, which is similar to the T1 in design, but provides for transmission at a 3.152 Mb/s rate in 22-gauge pulp-insulated cables and can provide 48 voicegrade channels (6.3 Mb/s), and it has a maximum range of 500 miles using special low-capacitance cable. The T4M transmits 4032 voice-grade channels (274 Mb/s) over coaxial cables. Although primarily designed for use in metropolitan areas, it has a range of 500 miles with repeater spacings of one mile (10).

There are three major digital radio systems in operation. The type 1A-RDS is a 1.5 Mb/s system used for the Digital Data System, often called DUV for "digital-under-voice," as it employs the previously unused frequency space beneath each analog radio channel. Another is the type 3A-RDS, an 11 MHz digital radio system, operating at approximately 45 Mb/s with repeaters spaced 6 to 25 miles depending upon terrain and weather conditions. This system was primarily designed for applications on high-density intercity routes up to 250 miles, as a feeder for the T4M or WT4 (274 Mb/s) systems and to provide route diversity within the network. The type DR18, an 18 MHz digital radio system, operates at 274 Mb/s with repeater spacings from one to five miles. The physical arrangement of this system is a pole-mounted canister that contains plugin modules for

seven working channels and one protection channel. It uses four-phase differentially coherent phase-shift keying, which assists in providing immunity from interference between radio channels on a given route and nearby routes (10,14).

Even higher capacities can be handled by metallic waveguides operating at millimeter radio wave lengths. The type WR4 is a long haul (4000 mile) digital waveguide facility which is designed to carry 230,000 two-way PCM voice circuits, and operates between 40 and 110 GHz. Because of low transmission losses, repeater station spacing of about 25 miles is anticipated. However, in view of expected developments in fiberoptics and satellites, and waveguide alignment problems, it is doubtful this system will be installed.

5.6.2 Satellite Communications Systems

Little note was taken of the earth satellite as a communications relay station until October, 1945 when Arthur C. Clarke published an article in an English technical journal which is considered to be the first public proposal of satellite communications. After the USSR launched Sputnik in 1957, the first true communications satellite, Score, rebroadcasted President Eisenhower's 1958 Christmas message. Several other low altitude satellites followed, including AT and T's Telstar, which was the first to transmit live TV across the Atlantic; and, in 1963, Arthur Clarke's prediction of a geostationary communications satellite came true with the successful launching of Syncom. The earlier

satellites had relatively low orbits (10,000 km) and passed quickly overhead. This required earth stations with elaborate tracking antennas, and to achieve continuous coverage between points on earth, a great many satellites would be required (50 to cover the North Atlantic). Higher orbits (38,500 km) and equatorial satellite placement made it possible, beginning with Syncom, to achieve a geostationary orbit. A single geostationary satellite can cover about four-tenths of the earth's surface and eliminates the need for tracking antennas. Since Syncom, about 80 geostationary satellites have been launched (72 built as communications satellites). Sixty-four of these were built and launched by the United States (57).

Five generations of Intelsat international communications satellites have been designed, built, and launched. These provide a history of geostationary satellite development. Intelsat 1 (1965) had a capacity of only 240 two-way telephone circuits and could link only a single pair of earth stations at any one time. Intelsat 2 (1967) was larger with increased power and bandwidth. Simultaneous access from multiple earth stations was permitted, but like Intelsat 1, the antenna rotated with the spinning satellite, directing a considerable portion of a radiated power out into space. Intelsat 3 (1968) used a despun antenna that always pointed toward earth (57). The entire allocated spectrum was used for 1200 two-way voice circuits. Intelsat 4 (1971) introduced

spot-beam transmitting antennas, concentrating the transmitted energy and permitting 4000 voice circuits. For the first time, capacity was limited by available bandwidth rather than available power. Intelsat 5 (1980) uses the 12/14 GHz bands and achieves a capacity of 12,500 voice circuits. Thus, in the 15 years of the Intelsat series, capacity has grown by more than 50 times; at the same time, the cost per circuit has dropped from $30,000 to $700.

While satellites have greatly increased their sophistication and capacity, earth stations have also become simpler. Antenna dimensions have dropped from over 70 m to less than 5 m. The cost of future stations could be less than $10,000 (58).

There are many challenges in expanding the satellite network in the future. For one thing, there are a limited number of orbit positions available. In addition, the 4/6 GHz arc is near capacity and the 12/14 GHz band is filling with applicants and the push to even higher frequencies is accelerating. At higher frequencies, however, new technology will have to be developed, and the absorption bands for H_2O (at 22.6 GHz) and O_2 at 60 and 200 GHz must be reckoned with.

Satellites are expected to become much more than simple repeaters, evolving to include sophisticated switching and satellite-satellite links as well as satellite-ground paths. Higher frequencies will permit more directional beams and satellites may well become space

platforms, equipped with scanning-beam or multiple-beam satellites. In a single orbital position, about 10,000 simultaneous wideband 3 Mb/s digital channels might be achieved (58). A summary of operational and proposed communications satellite systems is given in the references (18).

5.6.3 A Forecast of Communication Satellite Systems

This section presents quantitative forecasts of communication satellite systems in the United States. A substantial portion of material is based on results of a recent study (55). Generally, long-term projections are provided. A baseline scenario of communication satellite with a series of small incremental improvements in satellite characteristics is used. The cumulative effect of such improvements, nevertheless, offers a substantial increase in overall satellite communication systems capabilities. More specifically, trend projections cover satellite weight in orbit, number of satellites in orbit, average weight of satellite launched, primary power with and without sun-oriented arrays, satellite broadcast power, channels per satellite, total channels, and leasing costs per channel. The forecasts can be considered as a minimum expectation for communication satellites.

Communication Satellite Weight

The historical data and forecast trends of the total U.S. Communication Satellite System operational weight in orbit are shown in Figure 5.14. The 1975 weight of 9,773 kg is projected to

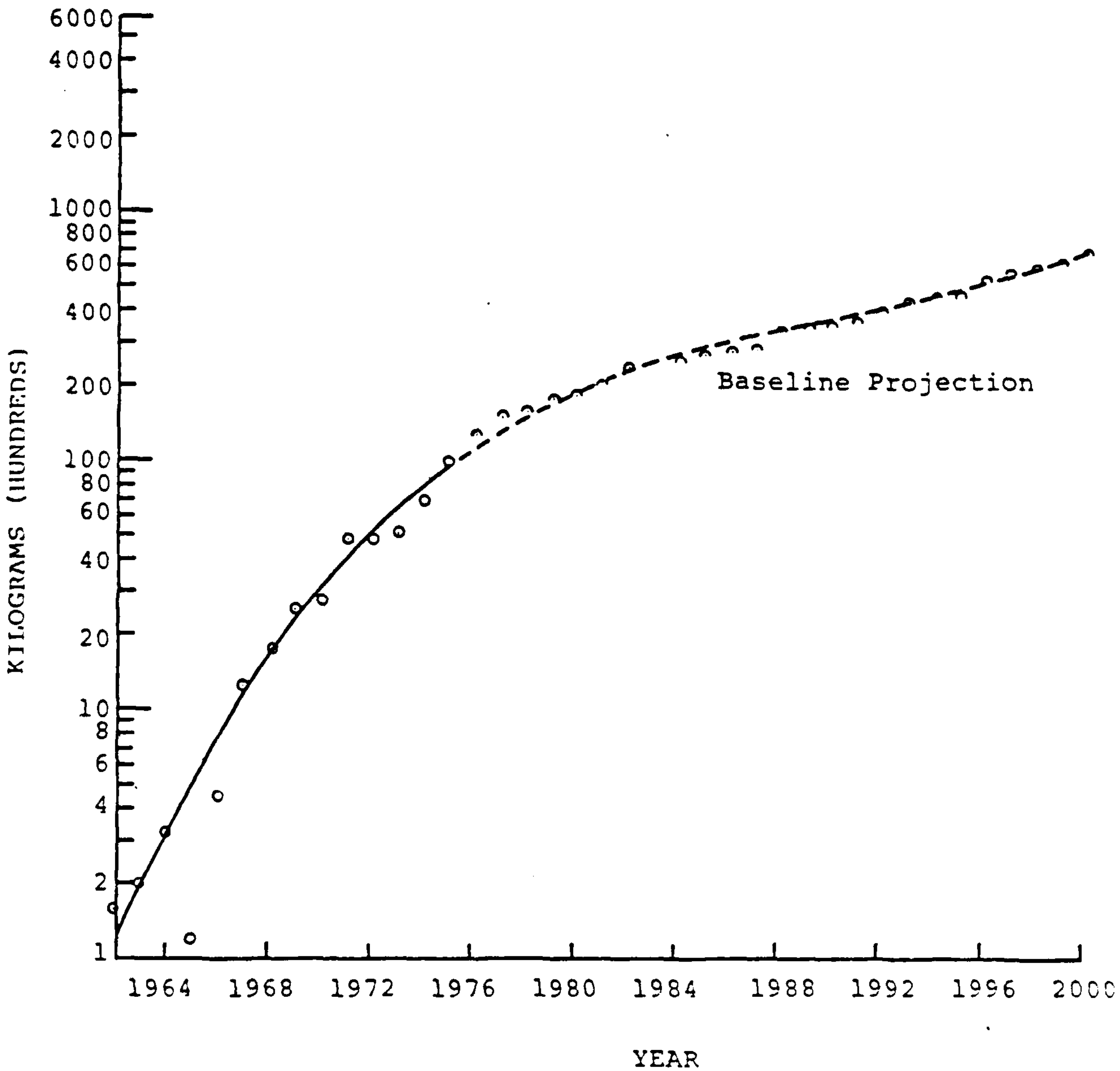
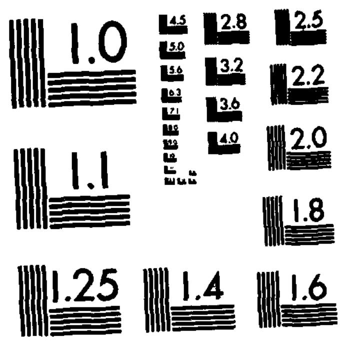


Fig. 5.14: U.S. Communication Satellite Weight in Orbit (Operational)

AD-A126 015 SOCIOECONOMIC IMPACT ASSESSMENT: COMMUNICATIONS INNUSTRY PHASE III TECHNO..(U) ACUMENICS RESEARCH AND TECHNOLOGY INC BETHESDA MD 02 FEB 79 FAA-APO-81-11-4 5/7

UNCLASSIFIED DOT-FA78WAI-932 F/G 17/2.1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

increase to approximately 67,600 kg by the year 2000, an average annual rate of 8 percent. This is conservative by comparison with the average annual rate of increase of nearly 40 percent per year during the period from 1962 to 1975, but is consistent with the recent pattern of a decreasing rate of increase. The perturbations in the projected curve result from assumptions regarding operating lifetime for each satellite, coupled with the known number of prior satellite launchings and the projected number of satellite launchings required to match with the projected number of satellites operational each year.

Number of Communication Satellites in Orbit

The number of operational U.S. communications satellites in orbit is shown in Figure 5.15. This number is expected to increase from 15 in 1975 to 78 in 2000. These figures include both military and commercial satellites. (For the Initial Defense Satellite Communication System (IDSCS), all satellite elements in each system launch are considered collectively as a single satellite in these calculations.) Since the rate of increase decreased during the period from 1967 to 1975 (after an initially very rapid rate of increase), the projection may appear slightly optimistic. However, the potential for expanded applications for communication satellites justifies this steady rate of increase. If the rate of increase should decline, the most likely effect (maintaining internal consistency) would be a corresponding decline in the rate

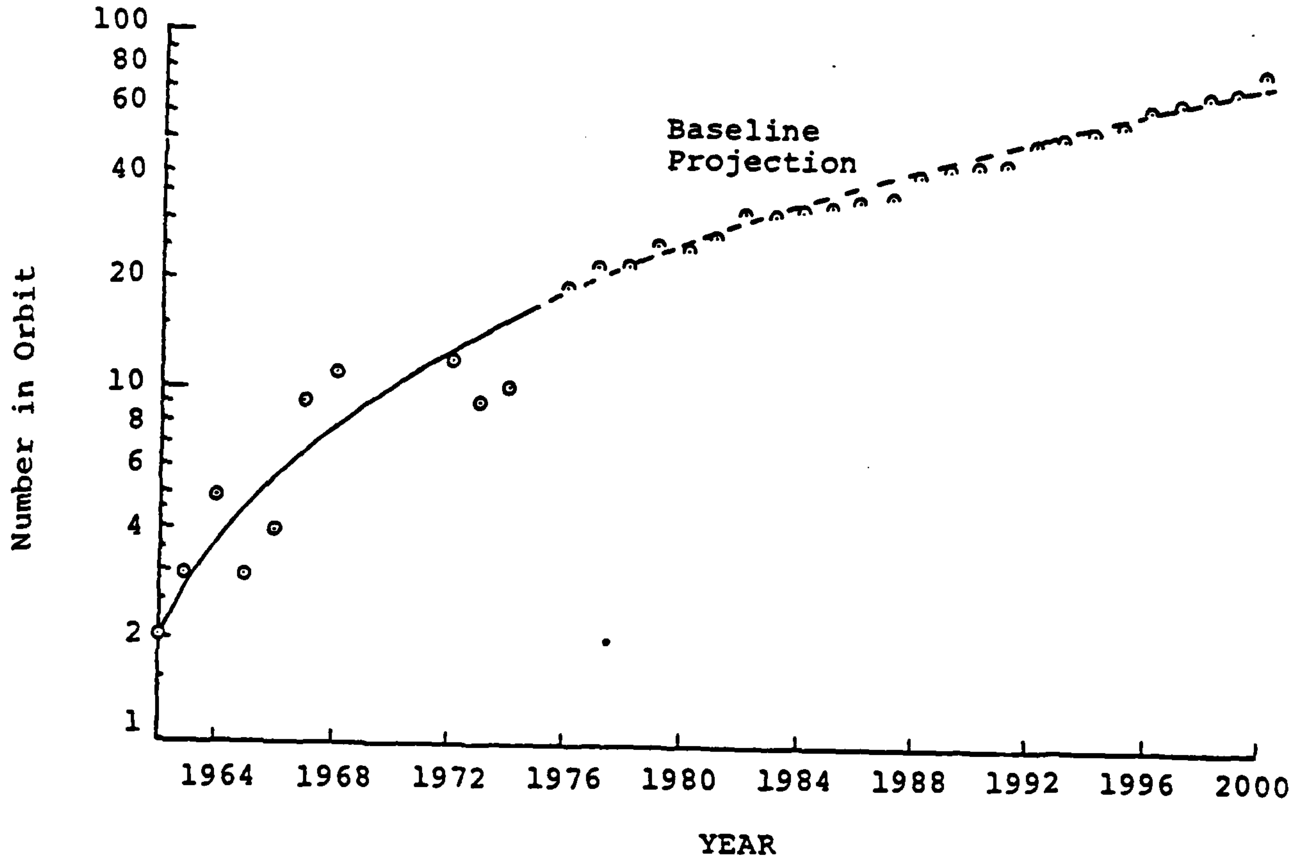


Fig. 5.15: Number of Operational U.S. Communication Satellites in Orbit at Year End

ACUMENICS

of increase of total communication satellite weight. Projections are made under the assumptions of no orbital slot allocation limitation, and additional frequency bands available, if needed.

Average Satellite Weight

Average operational satellite weight in orbit can be projected directly from the two preceding items (total operational weight divided by the number of operational satellites). This forecast is presented in Figure 5.16.

From the trend of total operational weight in orbit established in Figure 5.14, the total weight of communication satellites placed in orbit each year may be proejcted by adding to the increase in total weight each, the weight necessary to replace satellites which cease operations that year. The result of this calculation of the total weight of communication satellites to be launched each year is presented in Figure 5.16. From the launch weight total of 3,111 kg in 1975, the annual launch weight total is projected to more than double by 2000, to 7,120 kg, or an average annual increase of 3.4 percent.

The projection under the alternate scenario A is also included. The alternate scenario A follows a pattern consistent with prior development in communication satellite capability and remains within reasonable constraints of technical feasibility. The alternative projection is primarily characterized by two major step increases

ACUMENICS

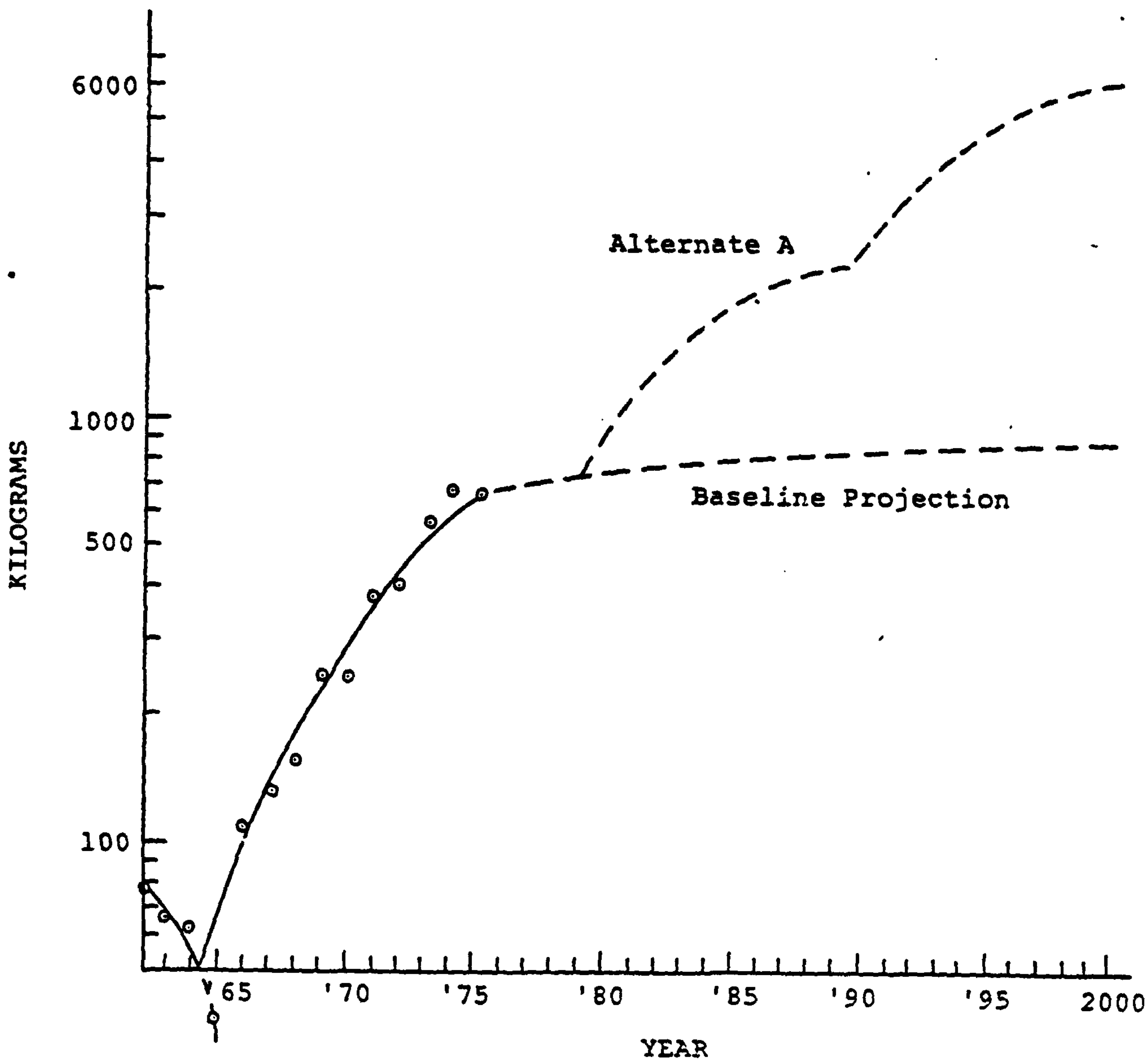


Fig. 5.16: Average Weight of Operational U.S. Communication Satellites in Orbit

in satellite size, coupled with the same increase in the number of satellites as forecast in the baseline scenario.

Although annual total launch weight is a derived calculation in this forecast in order to establish internal consistency, it could have been projected independently, although with less precision in view of the variation of this quantity during the historical period. The 3.4 percent rate projected is much lower than the average rate of 26 percent from 1962 to 1975, and is conservative in that sense. However, the 3.4 percent rate has been selected for the baseline projection to reflect a pattern common to many new technologies in which a rapid early rate of exploitation is followed by a period of slower steady increase. Variation in the trend of annual total satellite launch weight resulting either from an increase in the number of satellites launched or from an increase in average satellite weight, would, of course, increase total weight in orbit. A reasonably conceivable upper limit to annual total communication satellites launch weight would be a 16.4 percent annual rate of increase, shown by the curve identified as alternate scenario A (Figure 5.16). This would lead to an annual launch weight total of 48,000 kg in 2000, more than 15 times larger than the 1975 total of 3,111 kg. This would lead to a total wweght in orbit of 468,000 kg in 2000, which is consistent with a reasonably conceivable upper limit for total weight (it may be noted that this is nearly seven tiems the basic projection of total weight of 67,600 kg in 2000).

Satelite Primary Power

The primary source of power for satellites has been an array of silicon solar cells. No change to another primary power source (i.e., nonsolar) is projected for this forecast. The original projection of this trend, assuming no major change in the concept of satellite solar power arrays is shown in Figure 5.17. From this projection, using the numbers and weights of satellites previously projected, the average electric power per satellite and electric power per kilogram of orbital weight can be calculated as presented in Figure 5.19. It may be noted that the latter quantity shows a slight increase with time, a trend which could not be inferred directly from the historical data on power per kilogram of orbital weight for individual satellites.

The forecast, up to this point, fails to take into account the major change in solar array design concept, from arrays on the cylindrical surface of the satellite to deployable, flat, and continuously sun-oriented arrays. This change, already in use on the NASA-Lewis Communications Technology Satellite, and planned for next-generation commercial satellites, increases the efficiency of the solar collector system by at least a factor of three. Therefore, assuming that the newer form of solar array will be used on all communication satellites from 1980 on, the original baseline projection for average power of each satellite launched is multiplied by three in 1980 and thereafter. This revised baseline projection is shown in Figure 5.18 and is used to recalculate

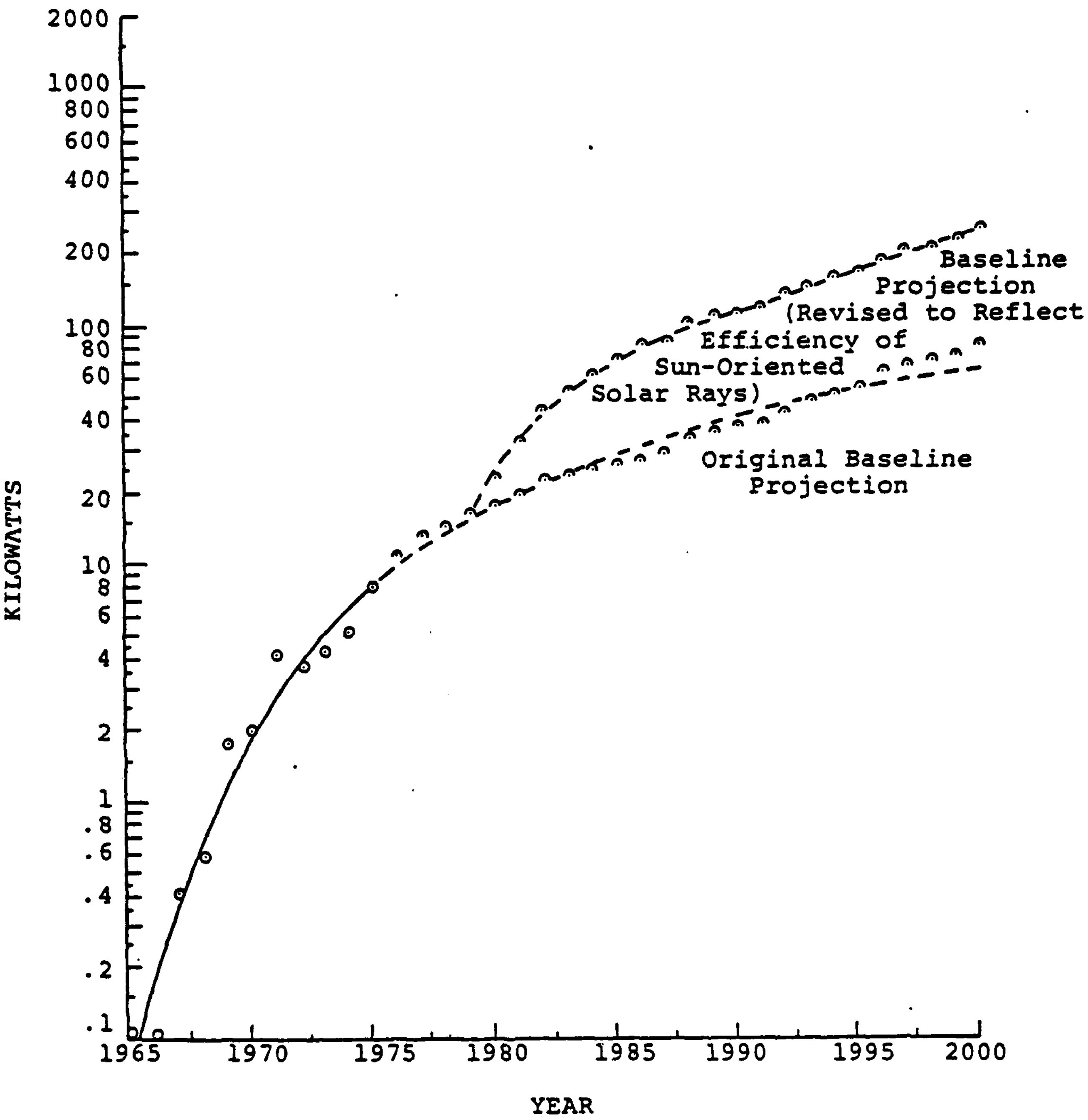


Fig.5.17: Total Electric Power of Operational U.S. Communication Satellites in Orbit

ACUMENICS

total power in orbit, average power in orbit, and total power launched each year, as plotted in Figures 5.17, 5.19 and 5.20.

Total operational satellite power (assuming sun-oriented solar arrays) is projected to increase at an average annual rate of nearly 15 percent during the next 25 years, reaching a total of 250 kilowatts in orbit in the year 2000. This is highly conservative compared to the rate of 55 percent per year for the period 1965 to 1975, but is consistent with the trend during the last few years.

Average operational power in orbit, plotted on Figure 5.19, shows the pattern which will prevail using sun-oriented solar cell arrays. The baseline projection shows the increase in power which will result if satellite size increases only marginally. Figure 5.20, showing the total power for satellites launched each year, dramatizes the sharp increase in power made available through the use of sunoriented arrays. Even with the smaller satellites of the baseline projection, more than ten kilowatts of new communications broadcast power would be added each year during the 1980s and two to three times that amount during the 1990s.

The average power of satellites launched each year should increase by a factor of six during the next 25 years, to 3.6 kilowatts per satellite, in the baseline projection. This would provide 30 kw of additional satellite broadcast power each year around the year 2000.

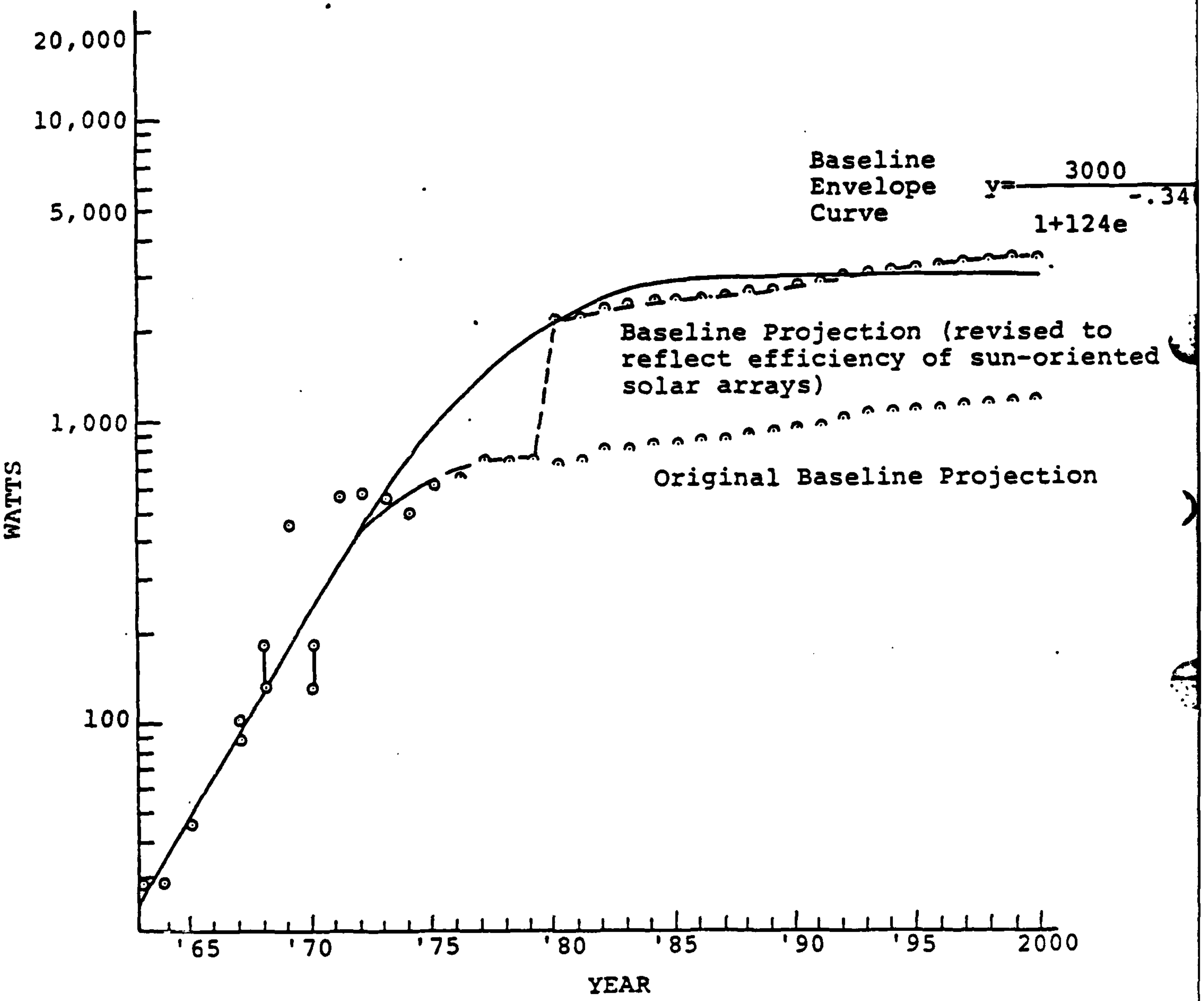


Fig. 5.18: Average Power of U.S. Communication Satellites Launched Each Year

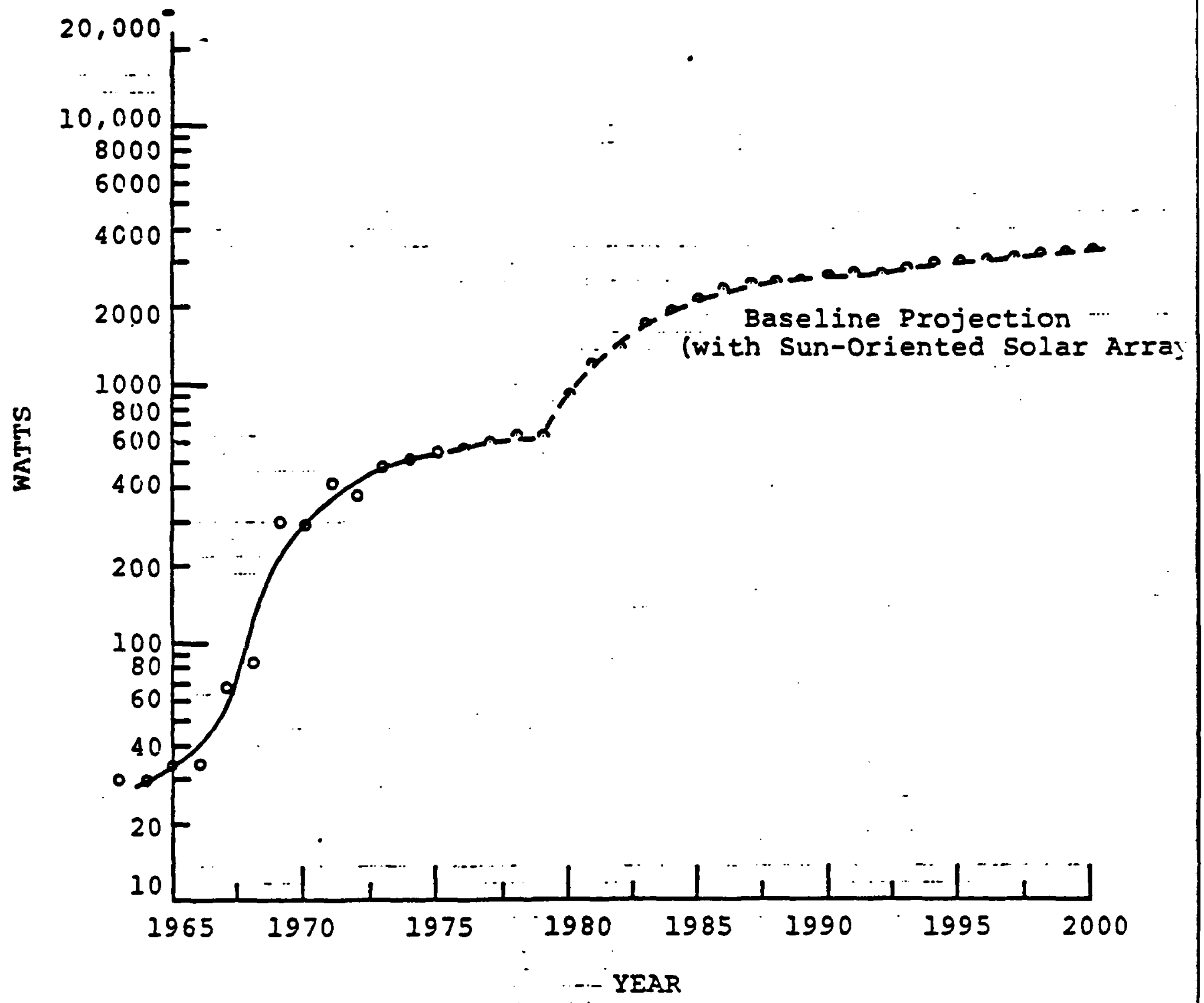


Fig. 5.19: Average Power of U.S. Communication Satellites in Orbit and Operational

ACUMENICS

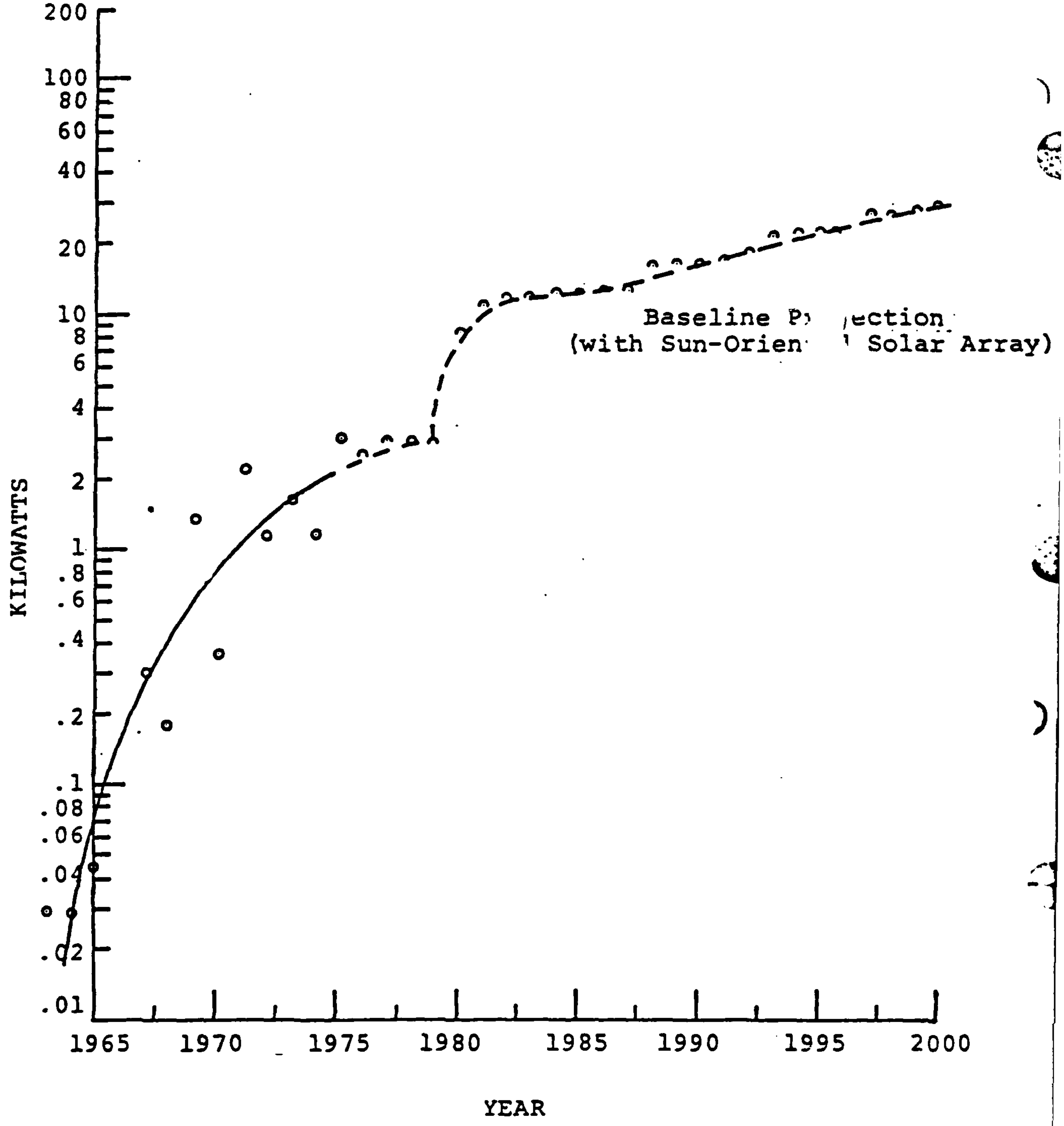


Fig. 5.20: Average Power of U.S. Communication Satellites Launched Each Year

ACUMENICS

Satellite Broadcast Power

The sum of Effective Isotropic Radiated Power (EIRP) per satellite is a measure of broadcast power and was chosen as a forecasting parameter primarily because it provides a reasonably uniform trend for projection. The regression of EIRP versus total satellite power availability is extrapolated and combined with the forecasts of total power to obtain projections of broadcast power. Granted that some fraction of total satellite power is used for other functions, it is assumed for this projection that such other uses constitute a relatively constant fraction of total power requirements. It is further assumed that the design for these uses will continue to minimize the fraction so that the projections of EIRP will be conservative.

Projection of EIRP using the Baseline is shown in Figure 5.21. Since EIRP for widebeam (17 percent) channels is limited to approximately 33 dBW for C-band (60-70 dBW in millimeter bands) by international agreements on terrestial flux density (60), and since satellite development is in the direction of multiple spot-beam antennas to concentrate satellite transmitted power over separate small areas of the earth's surface (61,62), the total of power will be used primiarly to increase the number of wide-band transponder channels per satellite. The trend of EIRP per channel for widebeam and spot-beam channels is shown in Figure 5.22. The Pearl Curve approach (63) was selected to project the trend of

EIRP per channel, because the previously sited references clearly imply growth of these characteristics toward upper limits. In the near future, C-Band is likely to dominate and therefore delay the rate of EIRP increase shown by the upper curve of Figure 5.22. On the other hand, the use of the 12 GHz frequency on Intelsat V, and a general trend in demand for higher EIRP, tend to support the spotbeam curve.

Number of Wide-band Channels Per Satellite

The Pearl Curve for the wide-beam transponder channels provides the basis for projection of the number of channels per satellite since the total radiated power in the spot-beam channels is equivalent to the radiated power of the wide-beam channels, given equivalence in other factors. The projected number of channels per satellite, derived from total EIRP and EIRP per wide-beam channel, is shown by Figure 5.23. The baseline projection will provide 36 to 48 channels (for simultaneous operation) per satellite in the period from 1980 to 2000. At 750 duplex voicegrade channels per channel, the baseline projection would provide 27,000 to 36,000 simultaneous voicegrade channels per satellite from 1980 to 2000.

Satellite Wide-band Channel Capacity

The projection of the average number of 6 MHz channels per satellite launched together with the prior projections of satellites launched and satellites phasing out enables a calculated projection of the number of wide-band channels which may be expected to be

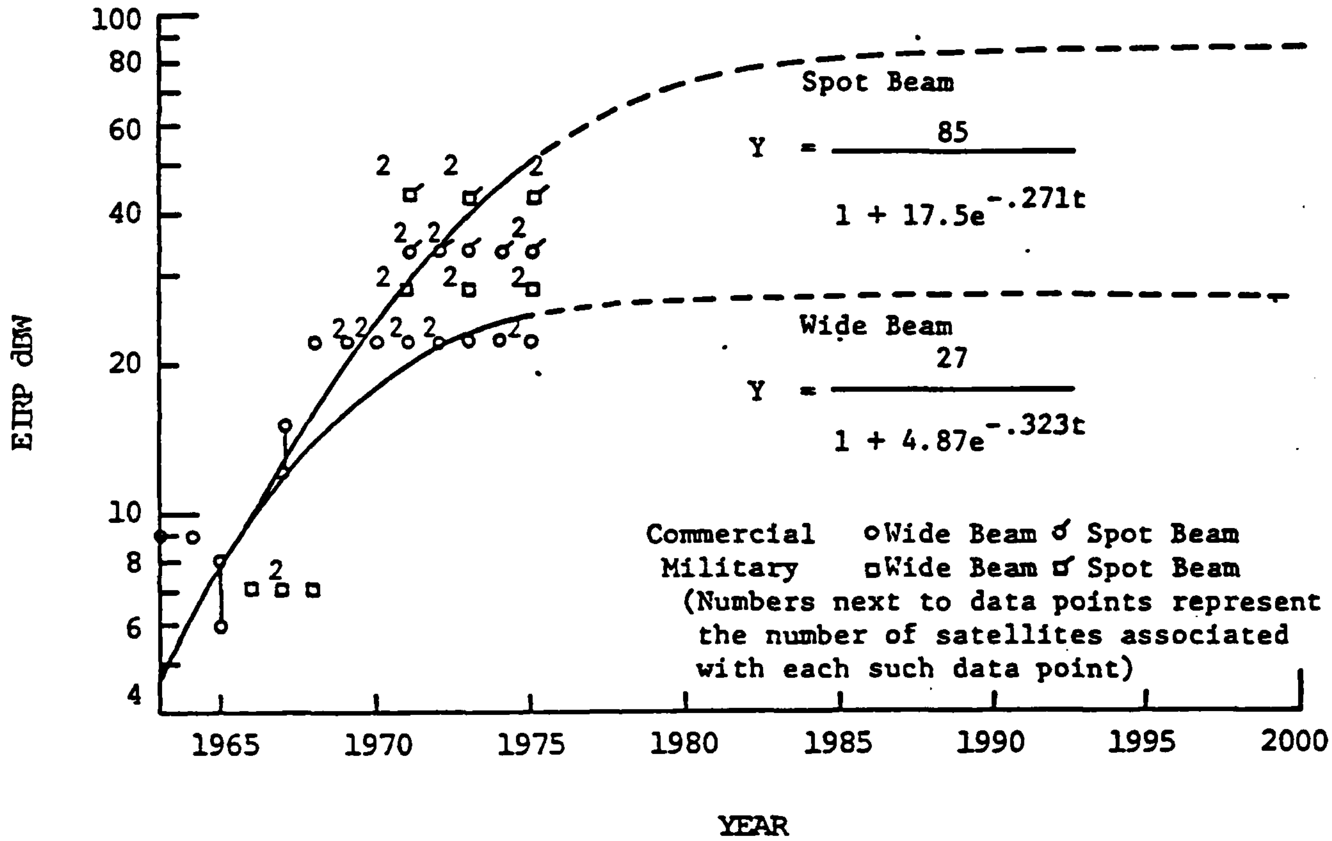


Figure 5.21 Effective Isotropic Radiated Power (EIRP) Per Channel

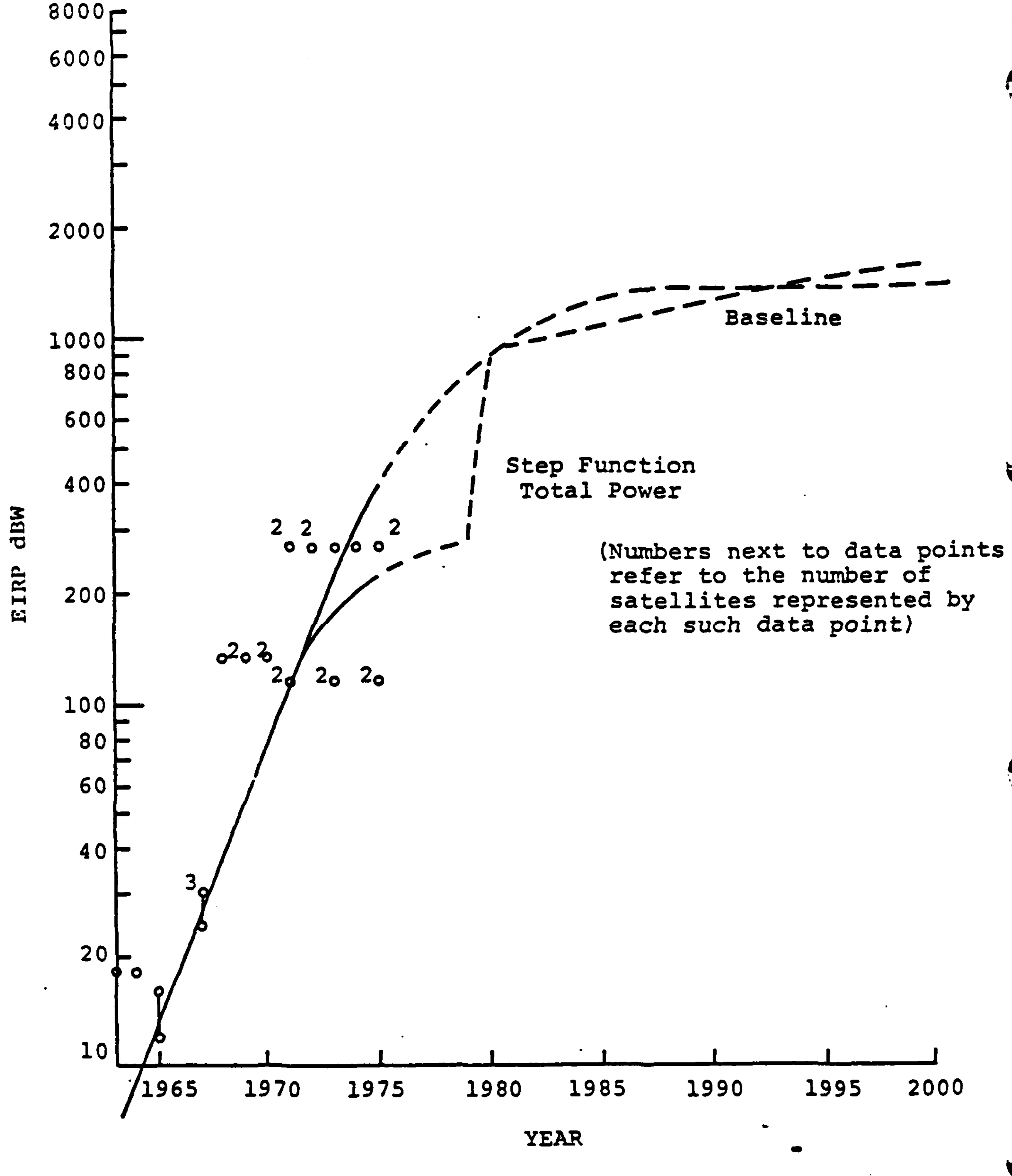


Figure 5.22 Total EIRP Per Satellite Launched
(Global Beam 17°)

ACUMENICS

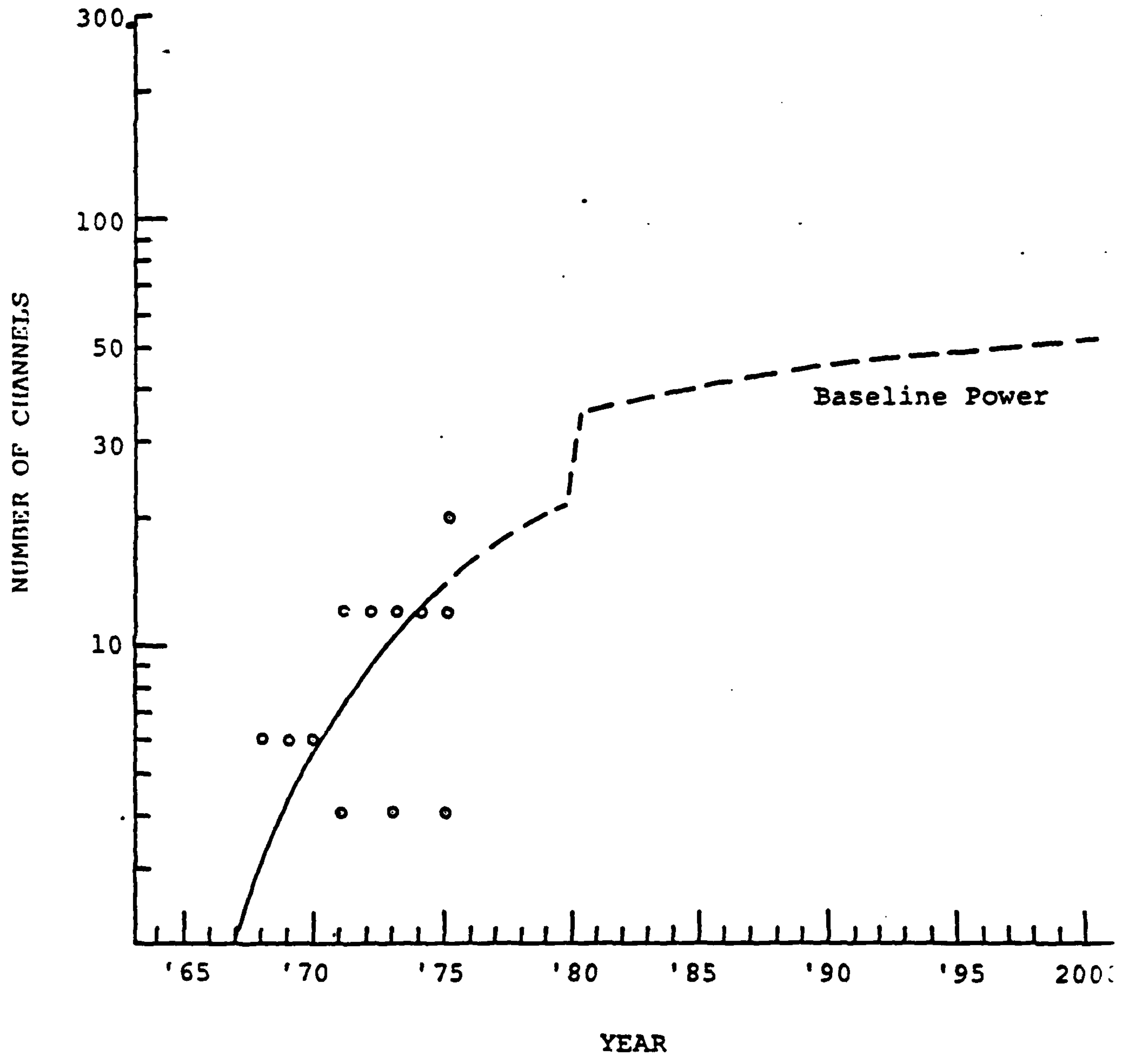


Figure 5.23 Number of Wideband Channels per Satellite

ACUMENICS

operational each year. For the baseline projection, this calculation indicates an increase from the 160 such channels in 1975, to 612 channels in 1980, 1310 in 1985, 2005 in 1990, and 3822 in the year 2000. The average annual rate of growth would be 14 percent starting at 23 percent in the early years and slowing to 7 percent near the end of the century; their rates are conservative in comparison with the 1965-1975 annual rate of growth of approximately 40 percent per year.

The projections, as shown in Figure 5.24, are based on the envelope curve of the power projeciton. If instead the number of channels is based on the assumption of step function increases in power as a result of the change to sun-oriented solar arrays and the two step increases in satellite size, then the end results (year 2000) are the same, but the 1980 and 1990 projections are reduced by 40-50 percent.

The baseline projection appears to stabilize at a 5 percent growth rate from 1985 to 2000. Thus, assuming that this growth rate will gradually slow, a further projection of the baseline may be made out to the year 2040. Using a 5 percent growth rate to 2010, 4 percent for 2011-2030, and 3 percent for 2031 to 2040, results in the following projection:

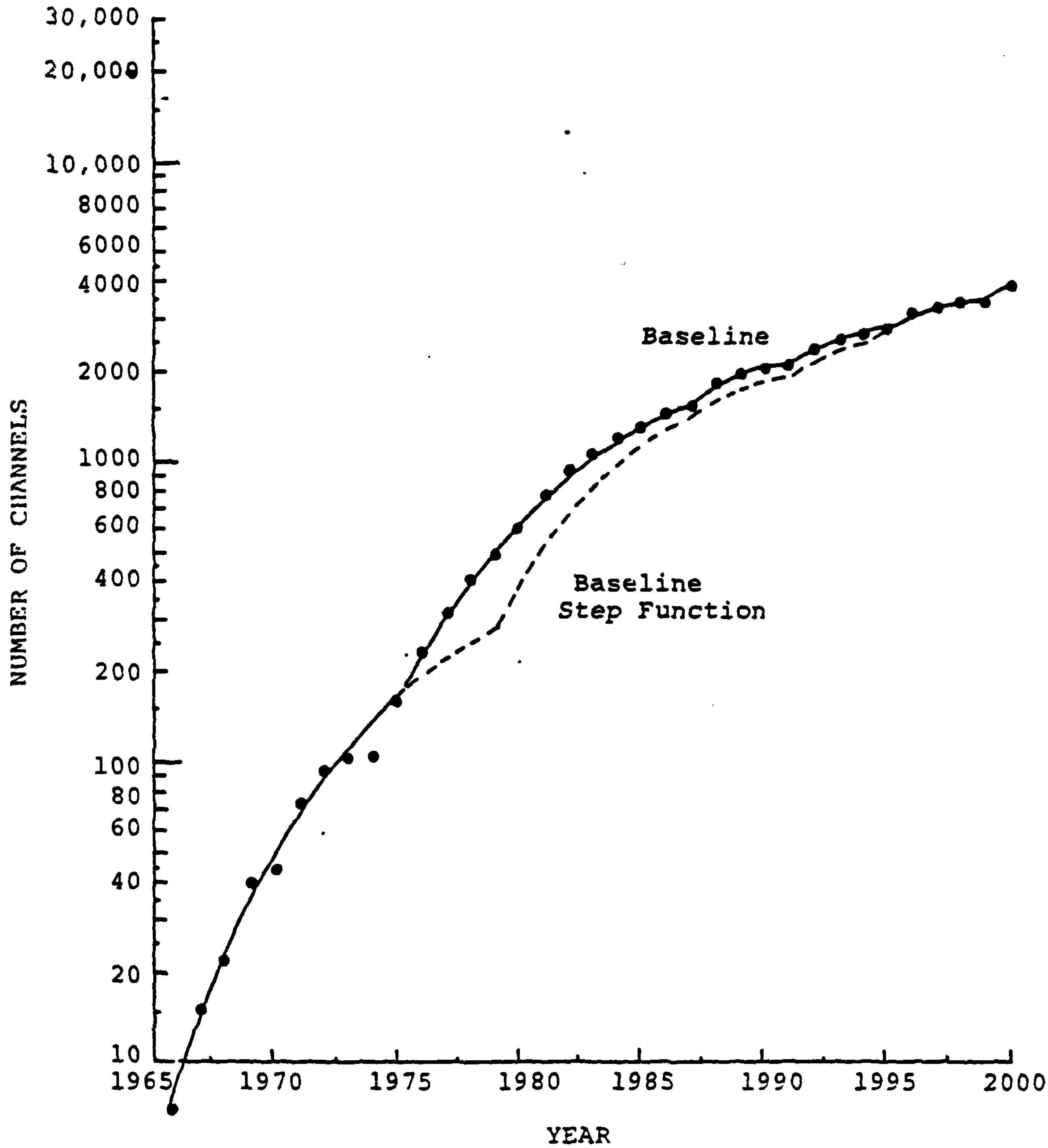


Fig. 5-24: Number of Wideband Transponder Channels in Operation

Year	No. of Channels	Year	No. of Channels
2000	3822	2021	9583
2001	4013	2033	9966
2002	4214	2023	10365
2003	4424	2024	10779
2004	4646	2025	11210
2005	4877	2026	11659
2006	5121	2027	12125
2007	5877	2028	12610
2008	5646	2029	13115
2009	5928	2030	13639
2010	6224	2031	14048
2011	6473	2032	14470
2012	6732	2033	14904
2013	7002	2034	15351
2014	7282	2035	15811
2015	7573	2036	16286
2016	7876	2037	16774
2017	8191	2038	17277
2018	8519	2039	17796
2019	8860	2040	18330
2020	9214		

Leasing Costs for Wide-Band Transponder Channels

A learning curve projeciton of leasing costs for wideband transponder channels is shown in Figure 5.25. The data for Intelsat leasing costs offers strong support for a 77 percent learning curve, which is reasonably consistent with the general experience factor in the electronics industry. Current data for various satellite systems indicates a wide scatter in wide-band transponder leasing costs, reflecting various pricing policies. There is evidence which indicates some cross-subsidization of full transponder costs by charges for primary-line circuits. This may result from the fact that leased-land-line service is currently high enough to provide a pricing margin for satellite transmission which permits a

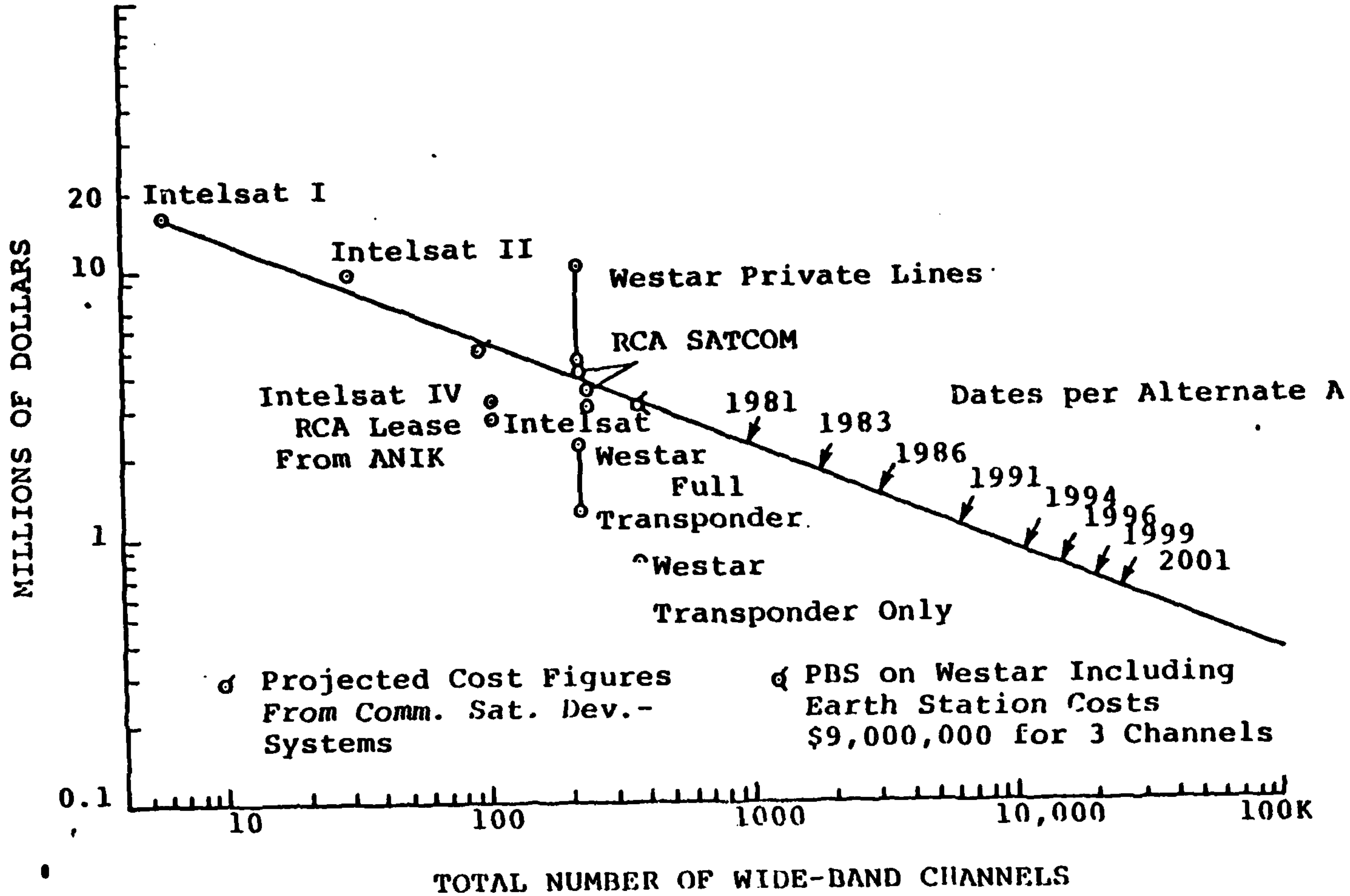


Figure 5.25 Learning C[illegible]e Projection of Lea[illegible]ing Cost Per Wideband Transponder Channel Ver[illegible] the [illegible]er of Operational

charge for satellite leased-line service which exceeds actual costs. The total evidence suggests that the 77 percent learning curve is conservative, i.e., that leasing costs may decrease at an even greater rate. The costs for transponder lease only, i.e., with earth terminal costs and distribution costs borne by the user, indicate an extremely rapid learning curve. However, the number of data points for full-transponder lease costs is insufficient to support any conclusion in this regard, other than the transponder costs, may constitute a decreasing fraction of the total communication costs.

Between 1980 and 2000, the leased cost per channel is forecast to drop by a factor of 50, with the number of channels increasing by at least an order of magnitude. This progress will facilitate a variety of new communication services, including mobile systems as described in the next section.

5.7 Mobile Telecommunication Systems

After years of development and testing, mobile telephone service was first introduced by the Bell System in 1946. A mobile user in those days had to select a channel manually, depress a push-to-talk button on the handset and place a call through a telephone operator. Today, mobile customers are able to dial their own calls on channels selected automatically. The need for mobile communications has outstripped available channels. The Bell System provides mobile service today to about 40,000 customers with about 20,000 more on the waiting list.

Mobile service is limited to a maximum of 25 channels and uses a single, relatively high-powered base station to serve an entire city. Channel reuse distances are in the order of 75 miles and, therefore, only a maximum of 25 channels are available in an area of 5000 square miles. This means that in New York City, as an example, only about 700 customers can be served and they experience a relatively poor grade of service because of the difficulty of finding an idle channel. The chance of completing a call on the first attempt is well under 50 percent.

In response to a 1968 FCC invitation for specific plans for innovative systems which could cope with the mobile system requirements, AT & T proposed a high-capacity cellular mobile concept. Testing of the system was initiated in 1978 in Chicago. The Chicago demonstration consists of 10 cells, each served by a low-power transmitter, a receiver and a control system--together making up a "cell site".

Every cell is allocated a set of frequencies with neighboring cells assigned different frequencies to avoid interference. Cells sufficiently far apart can simultaneously use the same frequencies--allowing for reuse of each channel for different conversations many times in a given service area.

Each cell will serve mobile customers when their vehicles are within that cell. As a vehicle moves from cell-to-cell, sophisticated electronic equipment hands off the call to another cell site.

site. In most cases, moving from one cell to another will not be noticed by the user. It is expected that the cellular mobile service will be extended to some 25 major urban areas by 1985 with the capability of handling hundreds of thousands of customers with an excellent grade of service (64).

More advanced satellite-aided mobile telephone systems are planned. These systems would augment the cellular-mobile terrestrial systems. The nation would be divided up into "footprints", each served by an earth station linked to one beam of a multiple-beam satellite. It has been estimted (65) that 69 footprints (beams) having a beamwidth of 0.5° could serve the entire continental United States. Using 333 channels of 10 MHz bandwidth, the system would permit 2850 busy-hour calls/footprint at a subscriber cost of about $25/month. Such systems are likely during the 1990s, extending mobile telephone service to millions of mobile subscribers.

5.8 Private Networks

Private networks provide communications service to customers with multiple locations, principally for internal communications. These networks are characterized by dedicated transmission facilities and some switching for use by a single customer. The chief motivation to use a private network is cost reduction. Private line costs per call drop as calling volume increases and beyond a certain calling volume, substantial cost savings are possible in contrast to the costs that would be incurred if the public switched network had

been used. Telephone companies are able to show that economic benefits arise when one considers the reduction of switching and other facility requirements in limited-size private networks. For example, the average call traverses 3.18 switches in the public switched network and only 1.67 in the typical private network (66).

Aside from ARINC, which runs the communications system for the domestic airlines, and the U.S. Government's Federal Telecommunications System, there are approximatley 750 firms which account for perhaps 80 percent of telecommunications spending in this country. These organizations, and perhaps another 4000-5000 medium size firms, make use of the services provided by AT and T's private line Tariff No. 260. The Tandem Tie Trunk Networks (TTTN) permitted by the tariff interconnect switches, located on the customer's premises, are tailored to meet individual requirements, being assembled from "piece parts" specified in the tariff. As TTTN networks make use of standard (older) telephone engineering and technology, few features are offered. There is no uniform numbering system nor a network transmission plan. Additional equipment is necessary to upgrade the older networks with traffic data recording, station message detail recording, and alternate route selection. Thousands of these networks are in use.

As a consequence of the introduction of stored-program control in switching, the introduction of competition in transmission

services, and increased customer sophistication, completely new network services have recently been introduced by AT and T. The Electronic Tandem Network (ETN) was introduced for use with the Dimension 2000 PBX and the No. 1 ESS Centrex, and has a TTTN-like network structure. Designed primarily for the medium-to-large customer, it consists of a two-level switching hierarchy in which the tandem network is fed by traffic from main or satellite PBXs. Software is provided that permits a uniform numbering plan, automatic alternate routing, queueing with off-hook ringback to let one know when a requested circuit is available, selective facility restriction, call detail and traffic recording and customer administration and control of the network (66).

The Enhanced Private Switched Communications Service (EPSCS) network concept was recently introduced for use by very large customers. It utilizes shared No. 1 ESS switches in the tandem mode which are interconnected CCIS-style through Peripheral Data Storage Processors which have their own data networks. The system makes use of four-wire switching which provides for improved voice and data transmission, and it permits station-to-station dialing using a uniform numbering plan. A conferencing feature permits any six stations in the network to be connnected. Automatic alternate routing is provided with economic route selection for calls bound for the public network. Three network-wide routing patterns can be activated at appropriate times of the day to accommodate changing

traffic loads. Call attempts and usage are summarized every half hour; the control center receives a count of the busy facilities every 100 seconds. Duplex data transmission at speeds up to 4800 bits per second is provided by the same access lines and trunks as the voice network (67). This new service requires the customer to order and maintain a minimum of 700 terminals and 125 terminals at each switch (68).

AT and T has implemented a digital network, specifically for data transmission, called Digital Dataphone Service (DDS). DDS offers a full duplex, point-to-point private line facility permitting synchronous data transmission speeds of 2400, 4800, 9600, and 56,000 bits per second. A 1.544 Mb/s speed was added in 1976. This service is available between more than 50 cities, and will eventually include 96 cities.

Packet switching is the outgrowth of the need to interconnect computers and data terminals. This concept permits many low and moderate volume users to share the same communication channel, thereby providing substantial cost savings for each user over the costs of conventional circuit-switched communications channels.

In late 1969, Bolt, Beranek and Newman, Inc., (BBN) placed into operation the first packet switched network in the U.S. called ARPANET. The ARPANET system architecture made use of minicomputers which were located at each network node and were interconnected in

a fully distributed manner by 50 Kbs leased lines. Each minicomputer took blocks of data from terminals and computers that were connected to it, subdivided these blocks into 128 byte packages (packets), and added a header which specified the destination; then, the minicomputer using a dynamically updated routing table sent the package over any free line which was currently the fastest route to the destination. Upon receiving the packet, the next minicomputer would acknowledge receipt and repeat the routing process independently. (69)

A user device or network-provided mechanism, in current packet switched networks, formats blocks of user data according to a standard set of conventions for transmission through the network. Two major elements make up a packet: the user's message and the network control information. Since the network is not transparent as with dedicated facilities, the network control information allows the user device and the network to maintain a liaison with each other.

The user data field is completely transparent to the network, and the user data can make use of any code. The user data can also be encrypted by the user if security is desired. Prefixing the user data field is a packet header. The three octet field specifies addressing information so that the network knows where to route the packet. It also contains packet sequencing numbers used for flow control. This assures that the sending data terminal equipment

(DTE) does not transmit data at an average rate greater than that at which the network or the receiving DTE can follow. These fields are packaged between frame delimiters containing an error detecting field and a link control field to insure the transmission across the access line into the network is accurate. High-level data link control (NDLC) Procedures, another part of international standards, specify the mechanisms of this link access protocol. The access link is full duplex to allow the transmission of packets in both directions simultaneously. Of course, the network is full duplex as well. Using the CCITT CRC-16 error checking polynomial, the entire user data field, packet header, and frame header are checked by the equipment at the node to assure that the network has received the same data as that sent by the originating station. At the receiving DTE, the same error checking procedures which occurred on the source line are applied to assure that the correct message has been received (74).

The first commercial authorization for a packet switched network occurred in 1973 when Packet Communications, Inc. (PCI) was permitted to lease data transmission circuits from established carriers. The PCI Decision and the Resale Decision of 1976 together solidified the position of PCI-like "Value Added Networks" (VANs) as viable contenders in the provision of telecommunications services. Data transmission using the packet switching concepts provides (75):

- Higher reliability -- through redundant circuitry, alternate routing and sophisticated error correction techniques.

- Compatibility for interfacing a wide variety of terminals and host computers.

- Economy -- frequently by an order of magnitude for low volume users over conventional circuit-switched services by efficient utilization of transmission media and distance-insensitive pricing.

- Flexibility -- in the choice of terminals, in adding new locations and applications, and handling growth of traffic.

Some idea of the growth of packet communications can be garnered from Telenet's experience which started in 1975 with seven network nodes. By middle 1978, the network had grown to 187 nodes, which provided 156 U.S. cities with local dial service to 180 host computers, with interconnections to 14 other countries (69). In 1979, the VANs (including Tymnet, Telenet, Graphnet and others) had an estimated revenue of $325 million with $425 million estimated for 1980 (68).

At the present time, microcomputers have been designed especially for packet switching and packet networks, and, as noted above, they are becoming universal translators, supplying needed speed, code and protocol conversion where necessary. Looking to the future, one major change that is forthcoming is the change from 56 Kbs to 1.544 Mbs (the speed of T1 digital trunks) in the backbone packet networks. With T1 trunks the transit time delay in packet networks will drop from the current 100-200 ms to 10-20 ms. This change to higher speeds will result in a significant decrease in line costs

to network users; however, a complete reexamination of network topology and processor design will be required. Such changes may make bulk data transfer attractive for processing through public packet networks (69).

Experiments with packet satellite techniques indicate that a single wideband channel (up to 60 Mbs) is an economical way to interconnect high bandwidth nodes with a packet network. It appears feasible to be used for batch traffic; however, the inherent 270 ms delay in satellite communication channels may be unacceptable for most interactive applications (69).

Digitized voice, no matter the digitization rate, can be compressed, by a factor as much as three, by packet switching (or by the Bell System's Time Assigned Speech Interpolation -TASI- methodology), since in normal conversation each speaker is speaking only one third of the time. Experiments with SATNET, an experimental packet-switched network for voice and data, operating at 64 Kbs, indicate the feasibility of satellite packet voice. Although the average host-to-host delay was found to be 1.4 sec with the majority of packets falling within 0.5 sec of this value, the speech experiments showed the relatively long delay from talker to listener did not cause significant problems. Improvement in hardware and software should reduce these delays substantially (70). Another area for improvement, previously mentioned, would be the increase in packet switching speeds to 1.544 Mbs or higher, which

would reduce packet delays to about 20 ms on coast-to-coast terrestrial circuits and close to 270 ms on satellite circuits, with the consequence that acceptable voice-data packet networks can be implemented (71).

Another future trend is the sharing of a wideband channel among many radio stations, each of which transmits in short bursts when it has data to send. Costs for this spread-spectrum application should become feasible in a very short time, considering the rapid cost reductions in semiconductor technology. Thus, such systems could be competitive with wire, coaxial cable or even fiber optics for low and medium volume local distribution requirements. Though the early ARPA packet radio experiments involved mobile installations, such low-power radio systems, when used in buildings, save the potential for eliminating the need for premises wiring for all types of telephone and data communications devices (69, 72, 73).

Western Union's Mailgram service was created as a joint venture with the U.S. Post Office in 1970, and has reached revenues now estimated to be $80 million per year and growing in excess of 20 percent per year. Mailgram revenues surpassed Western Union's telegram service in 1978 with a consequent result that competition in the public message computer-based services is starting to appear with Tymnet's On-Time Service having been tariffed and ITT and Telenet soon to offer competitive services (75).

New specialty communications services are expected to be offered soon by Satellite Business Systems, a joint venture of IBM, Comsat and Aetna Life Insurance. This will be the first domestic satellite system to employ the 12/14 GHz radio bands. The system will consist initially of two in-orbit satellites: one will be the primary operational satellite and the other will serve as a secondary operational satellite and backup for the primary. A third satellite will serve as a ground spare, to be launched if one of the in-orbit satellites fails. Use of the 12/14 GHz frequencies will permit smaller and less expensive earth satations to be located on the customer's premises. It is expected that either five or seven meter diameter antennas will be used, depending upon the customer's location. SBS estimates that by 1986 there will be 375 earth stations in operation. All traffic through the system--voice, data and image will be converted to digital format. Earth stations will transmit on a single-carrier per transponder basis, with several earth stations sharing the same transponder by time-division multiple-access techniques with demand assignment (TDMA/DA). A time division approach is much more efficient than the frequency division technique used by many existing systems. Time division multiplexing is illustrated in Figure 5.26 (76). With the SBS system, variale time slots will be assigned to earth stations depending upon traffic requirements. The user will order the number of access ports needed at each earth station. SBS anticipates that a user will have to lease a minimum of 16 voice-grade

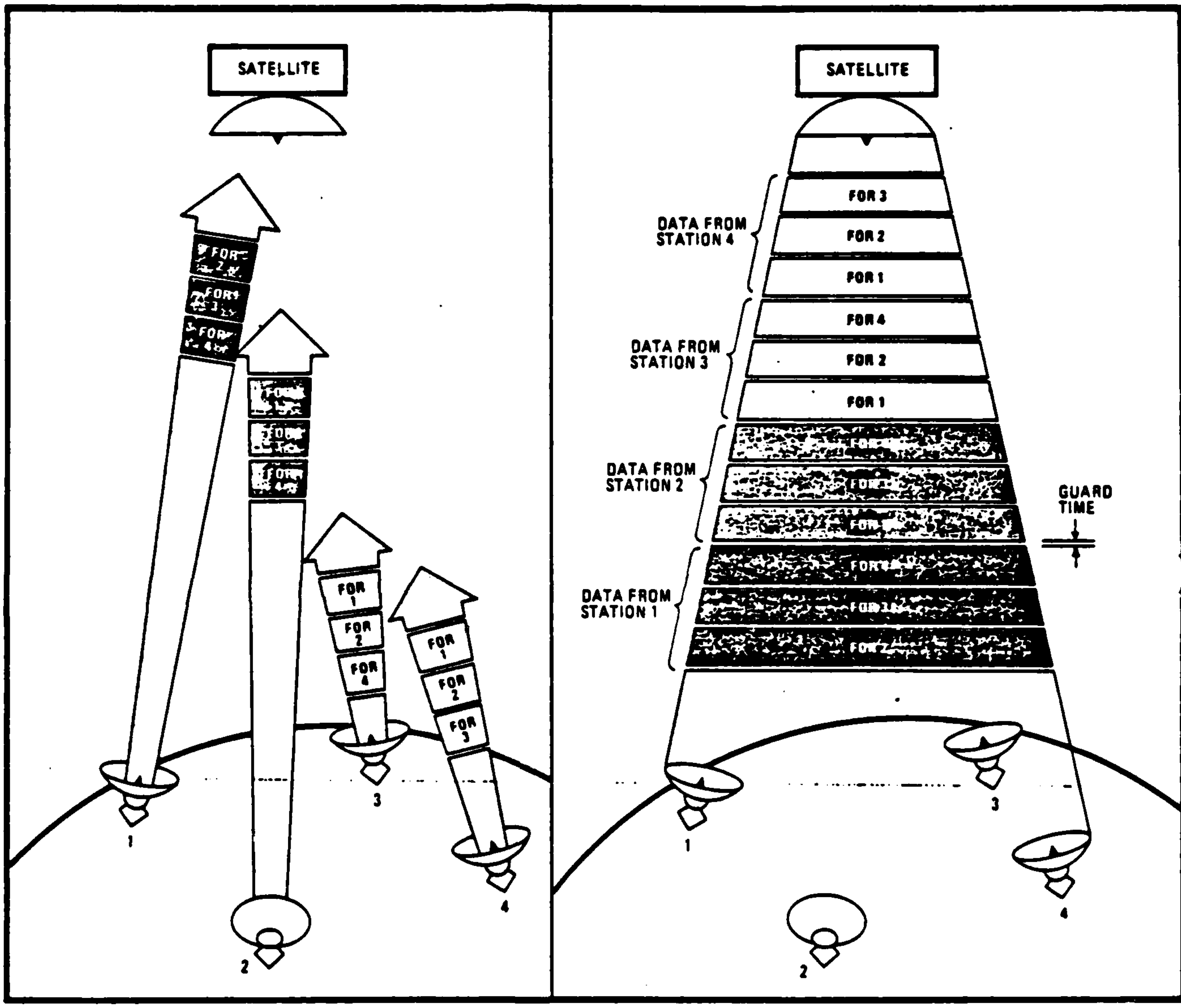


Fig. 5.26: Time Division Multiplexing.
The signal from an earth station is a set of bursts of digital data with coded addresses to particular receiving stations. Each burst has a synchronized time slot, so that interference is not a problem. Each receiver reads only properly addressed bursts. (76)

ports or one high-speed data port (56 Kb/s or greater) at each earth station.

Xerox Corporation recently received authorization to establish its proposed Electronic Message System (EMS), which would overcome the severe restrictions on data communications imposed by the design characteristics of existing local analog loops and switching equipment furnished by the telephone companies. The EMS concept would permit each subscriber to have access to a transmission rate of 256 Kb/s, compared with 9.6 Kb/s which can be derived from the local telephone company facilities.

The Xerox analysis of the potential market for the EMS service indicated that 100 MHz of spectrum would be required for local distribution. Such an allocation would be divided into ten channels, each of which would contain two 5 MHz subchannels, one for transmission and one for reception. This allocation would provide for up to ten nationwide systems, and Xerox has applied to operate one such network -- the Xerox Telecommunications Network (XTEN).

The XTEN network concept will offer a nationwide highspeed end-to-end digital communications service. The intercity backbone will utilize satellite transponders leased from current or prospposed domestic satellite carriers. Terrestrial point-to-point microwave systems may be used to carry traffic between the earth stations and the city nodes which, in turn, will use the new EMS spectrum to

communicate with stations located on the premises of individual subscribers. Each station transceiver will have a minimum transmission rate of 256 Kb/s. XTEN will employ cellular radio techniques to reuse the allocated EMS frequencies and to assure complete city coverage (similar to Bell's mobile radio cellular concept). The nationwide system will be under the control of two network control centers.

A cellular distribution of local nodes is planned -- each servicing a cell of approximately six miles in radius. Transmission is in the 10.565-10.615 GHz band. This makes available ten 5 MHz bandwidths channels for local distribution. Omni directional coverage from the city node is via either four 90° or three 120° sectorized antennas. Internodal control channels are used for frequency coordination and fault alarms.

The XTEN subscriber stations will transmit in the 10.630-10.680 GHz band at a 0.04 watt level and receiving will be accomplished in the 10.565-10.615 GHz band. Four-phase frequency shift keying will be used on 256 Kb/s channels. These stations will employ parabolic directional antennas directed toward a local node. The system will employ time division multiplex reception from a local node and time division multiple access transmission to a local node. The subscriber's processor provides store and forward service, priority control, security via encryption and password access and protocol matching to the subscriber's terminal equipment.

On July 10, 1978, A & T proposed its Advanced Communications System (ACS), forecasting the connection of 137,000 customer terminals and computers in 1983 out of an estimated 3.6 million terminals and computers expected to be in use then. In developing its ACS concept, AT and T noted that users have tended to develop a separate network for each purpose or application, resulting in compatible and under-utilized networks; users often have difficulty in modifiying existing networks to accommodate changing requirements; customers are finding that the tasks of managing multiple networks, including monitoring of their performance, is becoming an expensive and time consuming burden; and data communications systems often have startup costs that are too high to justify their use by smaller users or their deployment by larger users for applications that do not support substantial expenditures. The purpose of ACS is the computer-controlled transmission of messages wherein the substantive content of the message is not changed by the carrier. ACS includes a number of features that control and facilitate the transmission or movement of information or that are incidental to such transmission because they enable the customer to prepare, store and send messages in a more efficient manner. Interconnection between other networks and other carrier services will be possible through ACS network ports.

Although most of the traffic in both public and private networks involves voice, it is estimated that currently about 5 percent involves data traffic. Both public and private data networks

provide a mechanism whereby programs and resources residing in computers attached to any network may be accessed or altered. There are three basic types of networks (77):

1. Resource sharing networks provide the necessary connections to make it appear as if the remote sources (such as files) are available locally.

2. Distributed computation networks permit cooperating programs or processes in different computers to communicate and exchange information.

3. Remote communication networks permit remote interactive terminals to share communication facilities, by means of concentrators and multiplexers, in the movement of information to and from the host computer.

To design an appropriate network to meet specific requirements leads to the concept of "network architecture" so that, hopefully, the resulting network design can be flexible, allowing the interconnection of heterogeneous components, and adaptable, permitting the introduction of new technologies. Of these, probably the best known is IBM's System Network Architecture (SNA) which provides the structure for creating networks of IBM computers and components. End users of an SNA network are provided with a highly transparent, sequential, bit-stream channel, independent of topology, route selection or transmission media of the network (77).

As many new network architectures are being evolved, the connection of networks, or inter-networking, is of both national and international concern. A major force for change is the data transfer demands of multi-national corporations. In addition,

small networks and local or private networks will need to connect with other networks for growth and business purposes. In the future, rather than relying on massive monolithic computer networks, there will be a large number of public and private networks, each providing special services. Inter-networking, the technology by which autonomous networks are connected, is developing by necessity. Some of the technical issues include the need to translate from one network's protocol to another's, the need to define service, classes which can be mapped into each network's capabilities; the need to work out optimal routing and congestion control when interconnecting networks, and the management of multiple networks, including the identification and correction of service problems and the important function of apportionment of billing between networks.

Accordingly, interface standards are required to assure effective utilization of these new services. At the international level, the recommendations of CCITT Study Group VII and the International Organization for Standardization (ISO) are prominent. The American National Standards Institute and the Electronics Industries Association are active at the national level.

Examples of some important data network standards include (78) (a) the CCITT X.20 recommendations for data terminal and network interface requirements for stop-start services used with circuit-switched and leased circuits up to 300 bps; (b) the CCITT X.21

recommendations for synchronous operation at 600, 2400, 4800, 9600, and 48,000 bps. This standard represents a vast improvement over the older EIA RS-232-C. It provides a simpler electrical interface (much fewer connections), and greatly improved electrical characteristics. This standard has been adopted by the Scandinavian countries, the Federal Republic of Germany, and Japan. IBM indicates the use of the X.21 interface can functionally enhance its SNA networks (79); (c) the CCITT X.25 recommendations were specifically designed for data terminals connected to packet networks. The X.25 includes the X.21 recommendations for the physical, electrical, functional and procedural level to activate or deactivate the physical link between the data terminal equipment and the network; a link control level between the terminal equipment and the network to provide synchronization control, error deletion/correction functions for the transported information, and the network control level to establish end-to-end connections and the transfer of data through the network; and (d) the CCITT X.75 interface protocol is designed for the gateways between packet networks handling network information exchange, such as routing information, accounting information and the like (80).

5.9 Summary

The FAA's ATC telecommunications modernizaton program, outlined in Phase I of this report, leads to the deployment of a new, flexible data communications network, the National Airspace Data

Interexchange Network (NADIN), the two integrated ground-air and ground-ground voice communications systems -- the Small Voice Switching System (SVSS) and the larger Voice Switching and Control System (VSCS). These new systems will be state-of-the-art, making use of modern switching and solid state technology with exceptional flexibility for accommodating future growth and service requirements. These systems are to go into service in the 1980's and will function in an unprecedented, fast moving technological environment.

In this new environment, projections indicate not only a staggering increase in voice and data traffic but also the development of new services -- specifically, electronic mail and teleconferencing.* By the year 2000, electronic mail traffic could reach 25 percent of the total but dropping off as video services start to expand. Much of the video growth is based on video conferencing, which could grow to an estimated 53 percent of the total traffic by the year 2000. Some estimates indicate that video conferencing has the potential of reducing air travel by a factor ranging up to 8 percent by the year 2000.(81)

The same study estimates that there will be a saturation of satellite circuits by the early 1990's and that satellites will provide 2 percent of the voice traffic in 1980 and 25 percent by

---

*AT&T estimates the total annual calls on the public switched network in the year 2000 will reach 33 billion in contrast to 0.5 billion in 1949. MCI estimates the annual growth rate of local telephone calls to be 5 percent; toll and long distance, 15 percent; and international, 22 percent.

the year 2000; and 1 percent of the data traffic in 1980 and 60 percent by the year 2000. The new AT&T fiber optic submarine cable should alleviate future North Atlantic traffic demands somewhat as it will be in operation in the late 1980's with 12,000 digital circuits. With its new speech interpolation process, the equivalent of 36,000 voice circuits could be available on this cable.

The telephone utilities in the advanced countries throughout the world are moving rapidly to the integration of digital switching and digital transmission technologies to form the Integrated Digital Network (IDN) for voice service. AT&T's proposed Advance Communications Service is an IDN; while Satellite Business Systems proposed network is an IDN that will be completely independent of other networks.

The U.S. network will grow more digital in time, particularly with the introduction of the T4 (274 Mbs) fiber optic trunks in the late 1980's; however, much of AT&T's analog transmission plant will remain in service for at least another decade as the result of cost analyses showing the benefits of adding TASI-E speech interpolation equipment which has the effect of doubling analog circuit capacity; by loading in 300 more FM channels on many existing analog circuits; and, by the use of the recently introduced type AR6A single sideband system which more than triples the capacity of a typical type TH radio channel (6000 voice circuits vs. 1880).

Many of the independent telephone companies are rapidly converting their electromechanical central offices to digital -- some expect that close to half will be converted by 1985, and the Bell System will embark on its Class 5 office conversions to digital in 1981. (82,83)

A major obstacle to creating IDNs independent of the public switched network is the subscriber loop. Xerox's XTEN proposal would use a cellular microwave concept operating at approximately 10 GHz, connecting subscribers to satellite long haul circuits, while the SBS proposal would use satellite transmission direct to its customers.

The telephone companies believe that optical fiber loops connecting central offices to subscribers will not be used to any extent before 1985. Some experiments are underway in which optical fibers are run from a digital central office and terminated at remote concentrators, subscriber carrier equipment or PBXs as close to the subscriber as is cost effective. The remaining distance to the subscriber's terminal is spanned by digital transmission over metallic loops. One candidate for digital transmission over wire pairs is called "ping pong."

It is somewhat similar to time division multiplex, but only two points are involved and the two ends of the loop take turns in sending bursts of pulses, which are buffered at the transmitters and receivers so that the resulting link appears to be four wires with continuous pulse streams in both directions (82).

When one or more non-voice services such as data, facsimile, or slow-scan TV are added to the IDN for voice, the mixed service becomes the Integrated Services Digital Network (ISDN). The capability to provide all of these, services and many others, on one common network has the potential for substantial economies of scale advantages over separate dedicated networks. ISDNs (initially limited in scope) are expected to appear in various countries before 1985.

Switching technology and terminal equipment design have a synergistic relationship with the semi-conductor industry. Microcomputers on a chip are a reality and by 1985, in high volume, their prices could be on the order of two dollars each. Accordingly, this development will have an enormous effect on terminal and switching applications, including speech/data compression applications, which hold the promise of substantially increasing communication channel efficiencies. A number of 16-bit microprocessors are now available, and in the 1980s, 32-bit machines should appear. Semi-conductor RAM memories, currently available in 64K-bit chips, are expected to approach the megabit level by the mid-1980s. Non-volatile megabit bubble memories have recently been announced and most likely will replace disk memories for small and medium-size data processing applications.

Assuming an adequate market to amortize development costs. VLSI circuitry shows the promise of extraordinary reduction in

equipment costs, weight, size, and power requirements for future telecommunications switches and terminal apparatus.

Software development costs can be a major expense in the development of a large switch -- often overshadowing the hardware costs by several times. However, much has been learned in recent years and there seems to be more reliance upon a relatively few software assembly languages and compilers for telecommunications applications. Early computer-based telecommunications systems employed assembly languages, but more recently high-level languages have been used to substantially reduce software development costs. Techniques such as microcoding are used to extend the assembly language instruction repertoire to accomodate specific telecommunications functions in firmware (83).

In regard to air-ground communications and navigation requirements, the 1979 WARC agreement has made available an additional one MHz of spectrum to the aviation community in 1990. Should the need develop, further channel splitting of the current 25 KHz channels is technicaly possible, even going to 3 or 4 KHz single sideband channels. Looking to future ATC requirements, it is quite possible that the scenario developed by the RTCA Special Committee 120 may warrant futher consideration in combination with something similar to the military's Joint Tactical Information Distribution System (JTIDS) (84, 85). The RTCA Committee noted that during the introductory phases of ATC control message automation, it is

anticipated that controllers will, be called upon, minimumally, to monitor many of the computer-generated automatic processes. As experience and confidence is gained and accumulated with the computer's performance and capability, the need for human monitoring or supervision of automatic processes will diminish.

The earliest automated computer-generated ATC air-ground communications mode envisioned could utilize computer-generated voice response to assist in system monitoring. Subsequent user acceptance of the automatic ATC communications concept can be expected to encourage increasing use of (flat panel) visual display/synthetic voice readouts in the aircraft cockpit. It may be anticipated that in areas of low traffic density, parts of the system would not be utilized. With the cost reductions expected in semiconductor technology, airborne ATC transponders, data links and displays could become affordable for use by General Aviation aircraft.

REFERENCES

1. J. L. Flanagan, M. R. Schroeder, B. S. Atal, R. E. Crochiere, N. S. Jayant, and J. M. Tribolet, "Speech Coding," IEEE Trans. Commun., 27, pp. 710-737, April 1979.

2. B. Gold, J. I. Raffel, T. Bially, "A Study of Future Directions for Low Rate Speech Processor Research, Development, and Implementation," M.I.T. Lincoln, Lab. Technical Note 1978-43, Nov. 15, 1978.

3. S. I. Campanella, "Digital Speech Interpolation," Comsat Technical Review, p.127, Spring 1976.

4. A. M. Molinari and F. C. Vagliani, "Bit Rate per Channel Halving in PCM Multiplexes by Speech Interpolation and Adaptive Quantization," IEEE Trans. Commun., COM-26, May 1978.

5. J. D. Oetting, "A Comparison of Modulation Techniques for Digital Radio," IEEE Trans on Commun., vol. COM-27, pp. 1752-1762, Dec. 1979.

6. I. M. Jacobs, "Practical Applications of Coding", IEEE Trans. on Information Theory, vol. IT-20, pp. 305-310, May 1979.

7. C. L. Shannon, "A Mathematical Theory of Communications," Bell System Tech. J., vol. 27, pp. 379-423, 623-656, 1948. (This reference is already cited in Chapter 5.)

8. B. Gold, "Digital Speech Networks," Proc. IEEE, 65, pp.1636-1658, December 1977.

9. The Radio Frequency Spectrum--United States Use and Management, Office of Telecommunications Policy, Aug. 1975.

10. "Initial Report on Civil Aviation Frequency Spectrum Requirements 1980-2000, RTCA, DO-165, May 27, 1976.

11. "Tables of Frequency Allocations and Other Extracts", Manual of Regulations and Procedures for Radio Frequency Management, Office of Telecommunications Policy, Jan. 1978.

12. "42nd Annual Report - Fiscal Year 1976," FCC, Washington, D. C.

13. "Inquiry Relative to Preparation for a General World Administrative Radio Conference," FCC Docket 20271, Dec. 28. 1978.

14. "Air-Ground Communications - Operational Considerations for 1980 and Beyond," RTCA Document DO-162, Oct. 17, 1975.

15. ICAO Meeting (preparation for WARC), Montreal, May 16-June 9, 1978.

16. "Spectrum Engineering--The Key to Progress," Joint Tech. Advisory Committee, IEEE-EIA, 1958.

17. "Communications Technology in the 10 to 40 GHz Band," M/A Com. Inc., April 1979.

18. C. E. Powell, "Industry Gears Up for Free-Space Communications," Optical Spectra, June 1979.

19. R. C. Dixon, "Spread-Spectrum Techniques in Civil Applications," Proc. IEEE Inter. Conf. on Cybernetics and Soc., Sept. 19-21, 1977.

20. R. C. Dixon, Spread-Spectrum Systems, John Wiley, N.Y., 1976.

21. "EMC Analysis of JTIDS in the 960-1215 MHz Band," Office of Telecommunications Policy, March 1978.

22. L. A. Berry and E. J. Haakinson, "Spectrum Efficiency for Multiple Independent Spread-Spectrum Land Mobile Systems," NTIA Report 78-11, November 1978.

23. T. Dvorak, "Compatibility of Spread-Spectrum Systems with Narrow Band FM Receivers in VHF Mobile Networks," IEEE Inter. Sym. on EMC, June 20-22, 1978.

24. W. Ginsberg, "Telecommunications and the U.S. Information Economy: 1985-1990," Paper prepared for Project Delta, McGill University, Monreal, Canada, June 14, 1979.

25. "Spread-Spectrum Techniques for the Space Shuttle," B. H. Batson and G. K. Heaath, 1979 National Telecommunications Conference Proceedings, November 27-29, 1979, Washington, D. C.

26. "The Use of Spread Spectrum Modulation for the Co-channel Transmission of Data and Television," D. C. Coll and N. Zervos, 1979 National Telecommunications Conference Proceedings, November 27-29, 1979, Washington, D. C.

27. H. Falk, "Communications: Technology Forecasting," IEEE Spectrum, pp. 42-45, April 1975.

28. H. Falk, "Picturephone and Beyond," IEEE Spectrum, pp. 45-49, November 1973.

29. "Loop Switching System Triples Cable Capacity," Bell Laboratories Record, May 1979.

30. "Subscriber Loop Carrier System," Bell Laboratories Record, May 1977.

31. "IBM Uses Infrared to Transmit Data," Optical Spectra, December 1979.

32. "A Field Unfolds in Free-Space: Infrared Data Transfer," Optical Spectra, January 1980.

33. R. Grundlach, "Fiber-Optic Developments Spark Worldwide Interest," Electronics, pp. 81-87, August 5, 1976.

34. K. C. Kao and G. A. Hockman, "Dielectric Fiber Surface Waveguide for Optical Frequencies," Proc. Inst. Elect. Engr. 113:1151-1158, London, July 1966.

35. J. Fagenbaum, "Optical Systems: A Review," IEEE Spectrum, p. 70, October 1979.

36. R. Gallawa, "U.S. and Canada: Initial Results with Fiberoptics," IEEE Spectrum, pp. 71-74, October 1979.

37. J. R. Jones, et.al., "Live Traffic Over AGT's Optical Fiber System," Telephone Engineer and Management, February 15, 1979.

38. S. Shimada, "Japan: Unusual Applications for Fiberoptics," IEEE Spectrum, pp. 74-76, October 1978.

39. J. E. Midwinter, "European Activity in Fiberoptics," IEEE Spectrum, pp. 76-77, October 1978.

40. I. Jacobs, "Optical Communication Technology for Networks," Proc. 1979 Conf. Comm. Networks, Jan. 30-Feb. 1, 1979, Washington, D. C.

41. J. N. Kessler, "Baseline for Fiber Optics Growth in Telephone Systems," Telephone Engineer and Management, September 1, 1978.

42. K. C. Kao and M. E. Collier, "Fiber Optic Systems in Future Telecommunications Networks," Telecommunications, April 1977.

43. I. Jacobs and S. E. Miller, "Optical Transmission of Voice and Data," IEEE Spectrum, pp. 33-41, February 1977.

44. T. Giallorenzi, "Optical Communications Research and Technology: Fiber Optics," Proc. of IEEE, 66:7,

45. T. Li, "Optical Fiber Communications - State of the Art," IEEE Trans. Communications, 26:7, July 1978.

46. F. L. Thiel and W. B. Bielawski, "Optical Waveguides Look Brighter than Ever," Electronics, pp. 89-96, March 21, 1974.

47. R. Love, "High-Performance Cables Achieve Zero Failure at Rated Tensile Strength," Electronics, pp. 88-90, August 5, 1976.

48. C. K. Kao and J. E. Goell, "Design Process for Fiber-Optic Systems," Electronics, pp. 113-116, September 16, 1976.

49. International Fiber Optics and Communications Handbook, Information Galekeepers, Brookline, Mass., 1979.

50. R. Goodfellow, A. C. Carter, I. Griffith, and R. R. Bradley, "GaInAsP/InP Fast, High Radiance LED's," IEEE Trans. Electron Devices, 26, pp. 1215-1219, August 1979.

51. E. G. Dierschke, "High Efficiency Double Heterostructure Hemispherical Infrared Emitting Diodes," IEEE Trans. Electron Devices, 26, pp. 1210-1214, August 1979.

52. D. Botez and M. Ettenberg, "Comparison of Surface- and Edge-Emitting LED's for use in Fiber-Optical Communications," IEEE Trans. Electron Devices, 26, pp. 1230-1237, August 1978.

53. F. D. King, "High-Radiance LED's have Linear Response to Analog Inputs," Electronics, pp. 92-94, August 5, 1976.

54. J. T. O'Brien, "Laser Diodes Provide High Power for High-Speed Communications Systems," Electronics, pp. 94-96, August 5, 1976.

55. J. P. Martino, R. C. Lenz, K. L. Chen, P, Kahut, W. Sekely, and J. Weiler, "A Cross Impact Methodology for the Assessment of U.S. Telecommunications Systems with Application to Fiber Optics Development," Executive Summary and Volumes I and II, Final Report prepared under Contract NAS3-20365 from NASA-Lewis Research Center, University of Dayton, Dayton, Ohio, 1979.

56. H. L VanTrees, E. V. Hoversten, and T. P. McGarty, "Communications Satellites: Looking to the 1980's," IEEE Spectrum, 14, pp. 43-51, December 1977.

57. P. L. Bargellini, "Commercial U.S. Satellites," IEEE Spectrum 16, pp. 30-37, October 1979.

58. B. B. Lusignan, J. G. Potter, and J. M. Janky, "Co-Op in the Sky: Public Service Satellite Consortium," IEEE Spectrum, 13, pp. 59-65, February 1976.

59. N. Mokhoff, "Satellite Communicatins," (three articles), IEEE Spectrum, 16, pp. 28-41, October 1979.

60. J. E. Kiegler, etal., "Momentum Wheel Three-Axis Atitude Control for Synchronous Communication Satellites," AIAA Progress in Astronautics and Aeronautics, Communication Satellite Developments: Systems, 33, MIT Press, Cambridge, Mass., 1974.

61. H. J. Meyerhoff, "Power Balancing in Multibeam Satellites," AIAA Progress in Astronautics and Aeronautics, Communication Satellite Developments: Systems, 41, MIT Press, Cambridge, Mass, 1976.

62. W. G. Schmidt, "Satellite-Switched TDMA: Transponder Switched or Beam Switched?," AIAA Progress in Astronautics and Aeronautics, Communication Satellite Developments: Systems, 41 MIT Press, Cambridge, Mass., 1976.

63. J. P. Martino, "Technological Forecasting for Decision Making," American Elsevier, New York, pp. 111-113, 1972.

64. J. S. Reed, "Cellular Plan on the Move in Chicago," Telephone Engineer and Management, May 15, 1978.

65. R. E. Anderson, R. T. Milton, and J. P. Brown, "Potential of Satellite-Aided Mobile Telephone," IEEE Trans. Comm., pp.73-81, January 1979.

66. J. H. Weber, "Private Networks," Telecommunications Technology Symposium, IEEE-NTIA, May 22, 1979, Washington, D. C.

67. S. S. Katz, etal., "A Sophisticated Switched Service," Bell Laboratories Record, February 1979.

68. H. Newton, "1979--The Emergence of a New Telephone Industry," Telephone Engineer and Management, January 15, 1979.

69. L. G. Roberts, "The Evolution of Packet Switching," Proceedings of the IEEE, November 1978.

70. W. Chu, M. Gerla, W. Naylor, "Experimental Results on the Packet Satellite Network," Conference Record of the 1979 National Telecommunications Conference, Washington, D. C. November 27-29, 1979.

71. H. Tanaka, et al, "Design of a Packet Voice Transmission System," Conference Record of the 1979 National Telecommunications Conference, Washington, D. C., November 27-29, 1979.

72. R. C. Dixon, "Spread-Spectrum Techniques in a Wireless Telephone System," Proceedings of 1977 International Conference on Cybernetics and Society, Washington, D. C., September 19-21, 1977.

73. R. E. Kahn, et al, "Advances in Packet Radio Technology," Proceedings of the IEEE, November 1978.

74. P. J. Dawling, "Packet Switching," Proc. 1979 Conf. on Communications Networks, Jan. 30-Feb. 1, 1979, Washington, D. C.

75. R. Gable, "VAN Services in the U.S.," Telecommunications, July 1978.

76. H. T. Hinden, "Transmitting Data by Satellite Calls for Special Signal Handling," Electronics, pp. 91-98, March 29, 1979.

77. Stuart Wecker, "Computer Network Architecture," Computer, September 1979.

78. H. C. Folts, "Status Report on New Standards for DTE/DCE Interface Protocols," Computer, September 1979.

79. V. Ahuja and S. A. Barclay, "Compatibility of System Network Architecture and the CCITT X.21 Interface," Conference Record of the 1979 National Telecommunications Conference, Washington, D. C., November 27-29, 1979.

80. V. G. Cerf and P. T. Kirstein, "Issues in Packet Network Interconnection," Proceedings IEEE, November 1978.

81. Symposium on Projected Telecommunications Traffic Trends, 1980 - 2000, National Research Council, Washington, D. C. Oct. 30, 1979.

82. R. G. DeWitt, "Trends in Common Carrier Technology at Continental Telephone," 1979 National Telecommunications Conference Proceedings, Nov. 27-29, 1979, Washington, D. C.

83. A. H. Solomon, "The Merging of Telecommunications and Information Processing," Computerworld, Jan. 21, 1980.

84. Air-Ground Communications - Operational Considerations for 1980 and Beyond, RTCA Document DO-162, Oct. 17, 1975.

85. EMC Analysis of JTIDS in the 960-1215 MHz Band, Office of Telecommunications Policy, March 1978.

INTRODUCTION TO THE CONSTRAINED FORECAST

The previous sections present a forecast of technology unencumbered with constraints imposed by economic or social factors. In this section, the technology forecast will be examined in the context of the three socio-economic scenarios defined previously. That is, the purpose of this section is to adjust the unencumbered forecast to reflect the constraints imposed by the scenarios. As such, it seems necessary to describe the relationship among the scenario variables and the technology defined in the forecast.

The present discussion required definition of the terms invention and innovation, as these two terms are often confused. Schnookler's taxonomy is the most direct in approaching this topic. Schnookler defines invention "simply as a prescription for a producable product or operable process so new as not to have been obvious to one skilled in the art at the time the idea was put forward...."[1] As such invention is part of the "social pool of knowledge of induced arts," i.e., technology.[2] "Technological knowledge may be used to produce either more knowledge or ordinary goods and services. A method of producing a given good or service is a technique. When an enterprise produces a good or service or uses a method or input that is new to it it makes a technical change.

---

[1]Jacob Schnookler, Invention and Economic Growth, Harvard, Cambridge, 1976.

[2]Ibid p. 6.

The first enterprise to make a given technical change is an innovator. Its action is innovation.

A such, invention is an alteration in the state of knowledge. Innovation is the act of bringing that knowledge into use. If innovation is an act of bringing technology to the market, then it is more often then not an act directed by entrepreneurs. As such, the factors which influence an entrepreneur's decisions, i.e. economic factors influence the rate of technological innovation. Whether a technology is introduced into a market in 10 years rather than 5 years is based on the economics of the market and the perceived risk associated with such economics by the innovator.

Each scenario considered in the present effort can be viewed as a different set of entrepreneur risk factors. As such, the timing of specific innovations will be a function of the scenario. That is to say, the technical knowledge necessary for invention x will be available under each scenario. As such invention x will become an innovation under each scenario. However the introduction of the innovation into the market will be influenced by the risk embodied in the economic environment. The innovation will be introduced quicker in an environment with less risk.

Adjusting the unconstrained technology forecast is viewed as weighing the risk of introduction, and altering the timing of innovation based on the assessment of risk. The measure considered an indication of risk is innovation.

6. Constrained Forecast

6.1 Introduction

The technology forecast in the previous section is based upon the project team's judgement with an implicit consideration of socio-economic phenomena. As such, the project team did not explicitly employ the socio-economic scenarios constructed in Phase II to bound or limit the forecast. The purpose of this section is to use the three Phase II scenarios to modify the technology forecast so that the results reflect the impact of different socio-economic phenomenon.

The three scenarios considered include: balanced growth, rapid growth and stagflation. A summary of each of the scenarios is presented below.

1. Balanced growth

   The rate and direction of technology and economic change have been explicitly constrained by society. The intent of regulating the nature and magnitude of technological growth is to minimize the untoward social and economic impacts derived from unconstrained growth. Industrial productivity increases at a slow but constant rate consistent with national policy. The population remains constant with the fertility rate being equal to replacement. Population migration results in increased growth in small cities, towns and rehabilitated inner cities. A decline

in population occurs in the suburbs of metropolitan areas. The traditional adversary relationship between government and industry shifts to a cooperative arrangement. Government regulation and economic competition are combined to attain specified social goals. The market place evolves into a complex of defined markets, stratified by being completely competitive or subject to significant regulation.

2. Rapid growth

National policy marginally influences the rate and direction of technological and economic change. In general, government does not act to control the rate and direction of social change. The development and diffusion of new technology is limited only by market forces. Industrial productivity increases rapidly to keep pace with the demand for intermediate and consumer goods. The national population increases with the fertility rate approaching post-World War II levels, and the replacement rate remains constant. Population migration is consistent with the patterns apparent during the 1960's - 1970's. That is, there is a net migration to the Southwest as well as continued suburban development in the East and in the Lower Great Lakes Region.

3. Stagflation

Attempts to formulate and implement a national policy for controlled economic and technological growth meet with uniform opposition from industrial interest. As such, the relationship between industry and government deteriorates. Government programs result in gerrymander regulations and laws that retard economic and technological growth. Industrial productivity decreases, unemployment increases, and inflation continues to diminish the currency. The lack of social progress increases social pressures to effect changes through increased government activity. One result of increased government activity is to reverse the deregulation trend initiated in the 1970's. Previously regulated communications industries are subject to new regulation in the public interest.

As previously noted, the unconstrained technology forecast was developed without explicit consideration of the three socio-economic scenarios outlined above. However, the project team did use the assumptions outlined in Chapter 1 to develop the unconstrained forecast. A comparison of the Chapter 1 assumptions and the socio-economic scenarios indicate that the unconstrained forecast posits are consistent with the balanced growth scenario. As such, it is reasonable to assert that the unconstrained forecast and the balanced growth technology forecast are equivalent. Therefore, the forecasted socio-economic parameters for the balanced growth scenario are attendant to the unconstrained technology forecast.

6.2 Constrained Forecast Method and Assumptions

The unconstrained technology forecast has been deemed consistent with the balanced growth scenario. However, the technology forecast, or rather pertinent parameters thereof, must also be cast in terms of the stagflation and rapid growth socio-economic scenarios.

The alteration of specific unconstrained technology forecast parameters to satisfy or be consistent with the remaining socio-economic scenarios requires that:

1. the assumptions or assertions be stated concerning the relationship between socio-economic phenomena and technological change;

2. the variety of socio-economic variables in each scenario be reduced to a minimal number of measures; and,

3. a reasonable means of qualitatively relating the three socio-economic scenarios be established.

6.3 Generic Assumptions

The following assumptions and/or assertions are the basis for the constrained forecast methodology.

1. The key assumption in a technology assessment is that socio-economic phenomena either influence or are impacted by technological change. This assumption while supported qualitatively cannot be proven quantitatively. Qualitative proof abounds in the current social environment. It is clear, for example, that aviation has altered the basis for intercity travel as well as the

number of people travelling, the conduct of business, and the location of population. While the preceeding changes are evident, by examination, a means has not been established to measure or ascribe a specific quantity of change to aviation.

2. A subsidiary assumption is that socio-economic phenomena induce changes in industrial productivity. Again, the assumption gains qualitative support when one reads industrial or economic history. For example, the phenomena in nineteenth century England which induced the shift to factory industry from cottage industry resulted in both technological and productivity changes.

3. A major assumption is that present and future technological change are not random. Rather, technological innovation is the result of planned or quasi-planned research and development activity. As such, future technological change will be evolutionary rather than revolutionary. Therefore, new technology or characteristics of technology derive from precursor equipment.

4. A corollary to the preceeding assumption is that given sufficient time the same technology would evolve irrespective of the socio-economic scenario. That is, different socio-economic scenarios will alter the rate but not content of innovation.

If the preceeding assumptions are accepted then one may assert that the relationship among scenarios can be expressed in terms of a numeric factor (technology acceleration factor). In similar, the changes in the technological forecast parameters will be a function of the numeric factor.

6.4 Technology Acceleration Factor Assumptions

The assumptions necessary to construct the technology acceleration factor (TAF) derive from the assertions delineated in Section 6.3. The assumptions for the TAF consider the relative relationship among the socio-economic scenarios. The assertions of import include:

1. The period between ideation and market introduction for a technology is influenced by the socio-economic scenario. It is likely that the time requirement, i.e. innovation lag, will be ordered as follows: stagflation, balanced growth, rapid growth. That is, it requires less time to move from ideation to market introduction under a rapid growth scenario than either balanced growth or stagflation.

2. Innovation diffusion or adoption times vary with particular socio-economic conditions. That is, the market is less likely to absorb new technology under a stagflation scenario than if rapid growth obtains.

3. Discrete historic parameters can be employed as surrogate measures for the proffered socio-economic scenarios. That is, the concepts embodied in the socio-economic scenarios have equivalents in economic history.

6.5 Technology Acceleration Factors

The calculation of the technology acceleration factor requires estimates of the innovation lag and diffusion times for technologies emanating from various industries. The data available concerning

average innovation lag times in four relevant industries is shown in Figure 6.1. Information concerning the diffusion of technology is more limited than innovation lag data. Heuristic guidance for diffusion is provided by the estimates of the time intervals between computer hardware generations shown in Figure 6.2. The innovation lag data in Figure 6.1 will be assumed to be consistent with the balanced growth scenario.

Innovation lag times for the balanced growth and stagflation scenarios will be estimated using historic indicators of per capita GNP. It will be assumed that 2000 represents a point estimate for stagflation, balanced growth and rapid growth. The data for GNP and population for 2000 are shown in Figure 6.3. The GNP/capita estimate scenarios are indicators also of productivity. If one makes the heroic assumption that innovation lag is a function of productivity, then the innovation lag for the rapid stagflation and balanced growth scenarios are a function of GNP/ capita for specific scenarios, or

$$L_i = L_j \quad \frac{Gj}{Gi}$$

where

L_j = innovation lag for balanced growth

L_i = innovation lag for stagflation or rapid growth

G_i = GNP/capita for stagflation or rapid growth

G_j = GNP/capita for balanced growth

The intermediate calculations of Gj/Gi and L_i are shown in Figure 6.4.

INNOVATION LAG

Industry	Average Time Lag Factor
Computer	5.4
Telephone	5.0
Communications	5.7
Electronics	3.3

Figure 6.1

DIFFUSION TIMES HARDWARE

Generation	Item	Years	Time Interval
0	Relays and Vacuum Tubes	Up to 1953	
1	Vacuum Tubes	1951-1958	7
2	Transisters	1958-1969	9
3	Integrated Circuits	1967-1974	7
4	Medium-scale and Large Scale Integration	1975-1979	4

Figure 6.2

Economic Indicators

Scenario	Point Estimate Year	GNP (1972 $) (Trillion)	Population Millions	GNP/Capita
Stagflation	2000	2.1	255.5	8,219
Balanced Growth	2000	2.7	262.5	10,286
Rapid Growth	2000	3.5	287.0	12,195

Figure 6.3

INTERMEDIATE CALCULATIONS FOR L_i

Scenario	G_j/G_i
Stagflation	$\frac{10,286}{8,219} = 1.25$
Balanced Growth	$\frac{10,286}{10,286} = 1$
Rapid Growth	$\frac{10,286}{12,195} = .84$

Figure 6.4

The estimates of innovation lags for the three scenarios are shown in Figure 6.5.

The technological acceleration factor is the ratio of a specific scenario and unconstrained forecast innovation lag. Therefore, TAF = $\frac{Gj}{Gi}$ in the present context.

6.6 Alteration of the Unconstrained Forecast

The purpose of this section is to delineate the techniques employed for modification of the unconstrained, i.e. balanced growth, technology forecast. The modifications will yield estimates of particular technology characteristics for both the stagflation and rapid growth scenarios. The unconstrained forecast values of specific technology characteristics for VLSI instrumentation system (Figure 6.6) will be used to demonstrate the technique.

The data in Figure 6.6 provides discrete values of certain characteristics for specified years. It is reasonable to assume that the data are points on a continuous function. A scatter plot of the data indicates that the function is not likely to be linear (Figure 6.7). There are not sufficient data points to analytically fit a curve to the data. However, the data are sufficient to allow a first order linear approximation between points (Figure 6.8). In addition, the data may be represented in polar form (Figure 6.9).

The estimation of technology parameters for the stagflation and rapid growth scenario will use data from both the linear

INNOVATION LAGS

	Innovation Lag Factor		
	Balanced Growth	Rapid Growth	Stagflation
Computer	5.4	3.5	11.53
Telephone	5.0	3.2	10.7
Communications	5.7	3.7	12.2
Electronics	3.3	2.1	7.1

Figure 6.5

VLSI INSTRUMENTATION SYSTEM
UNCONSTRAINED FORECAST

Year	Characteristic Power (W)	Characteristic Volume (in^3)	Cost ($1/1975)
1980	95	1,910	1,444
1985	77	805	901
1990	60	675	746
2000	55	590	535

Figure 6.6

ACUMENICS

VLSI INSTRUMENTATION SYSTEM POWER REQUIREMENTS
SCATTER PLOT

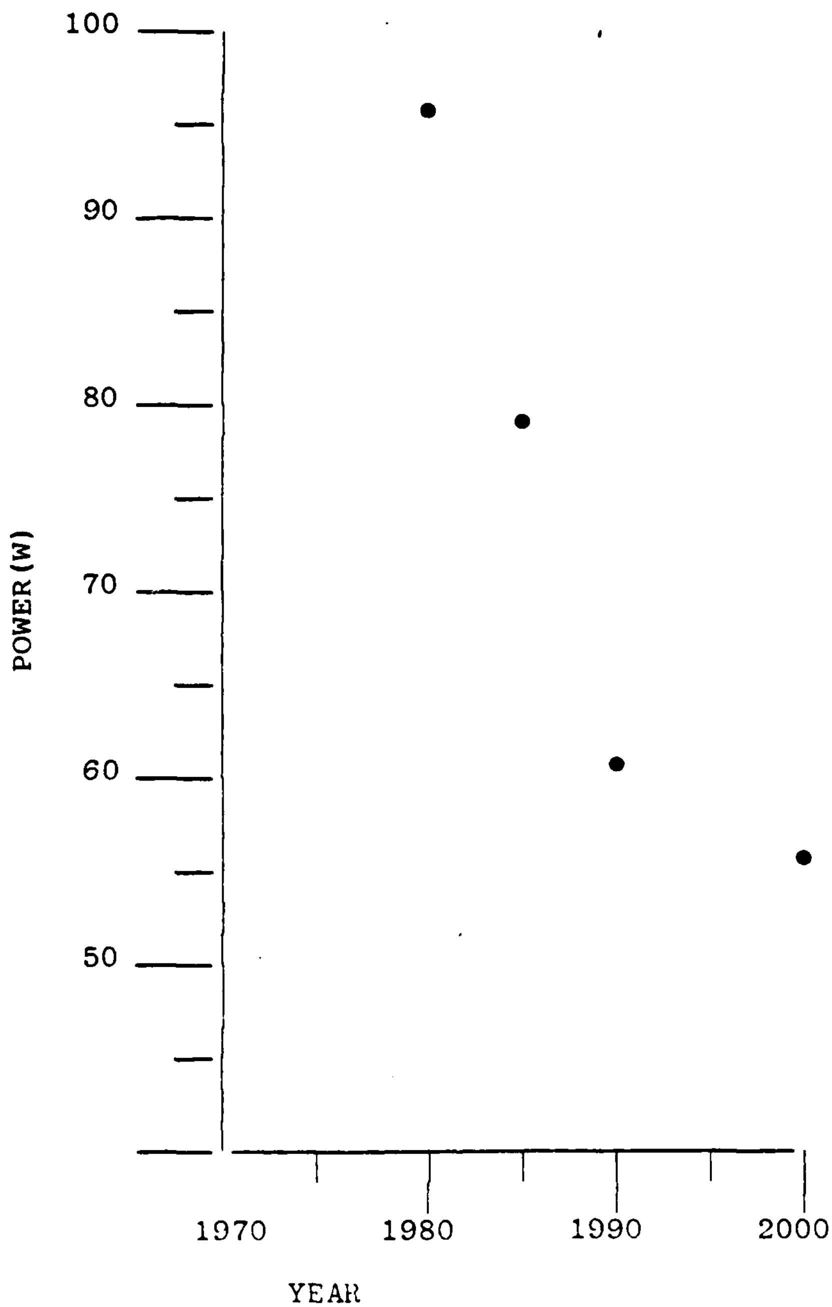


Figure 6.7

ACUMENICS

VLSI INSTRUMENTATION SYSTEM POWER REQUIREMENTS
LINEAR APPROXIMATION

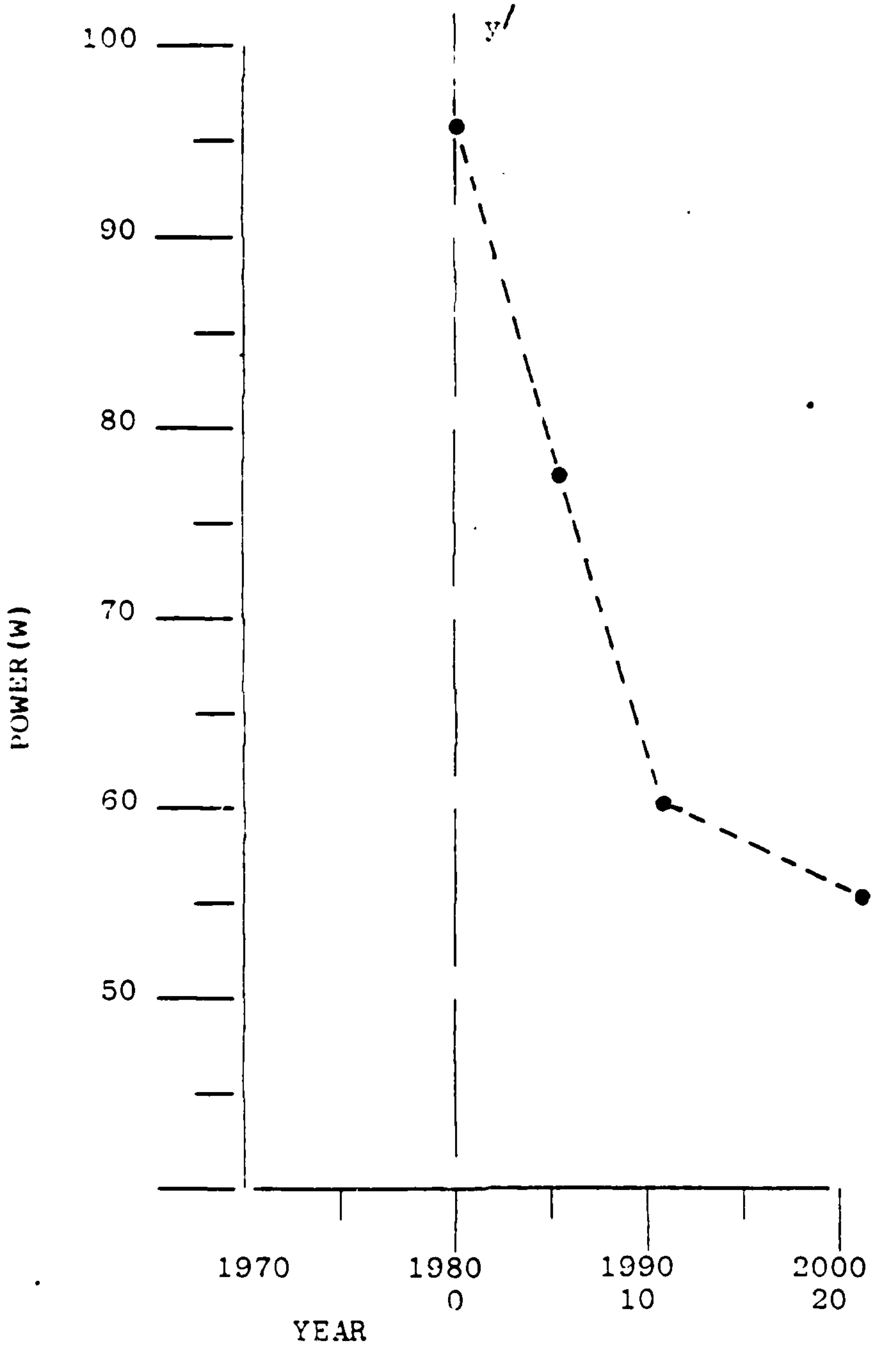


Figure 6.8

ACUMENICS

VLSI INSTRUMENTATION SYSTEM POWER REQUIREMENTS
POLAR REPRESENTATION

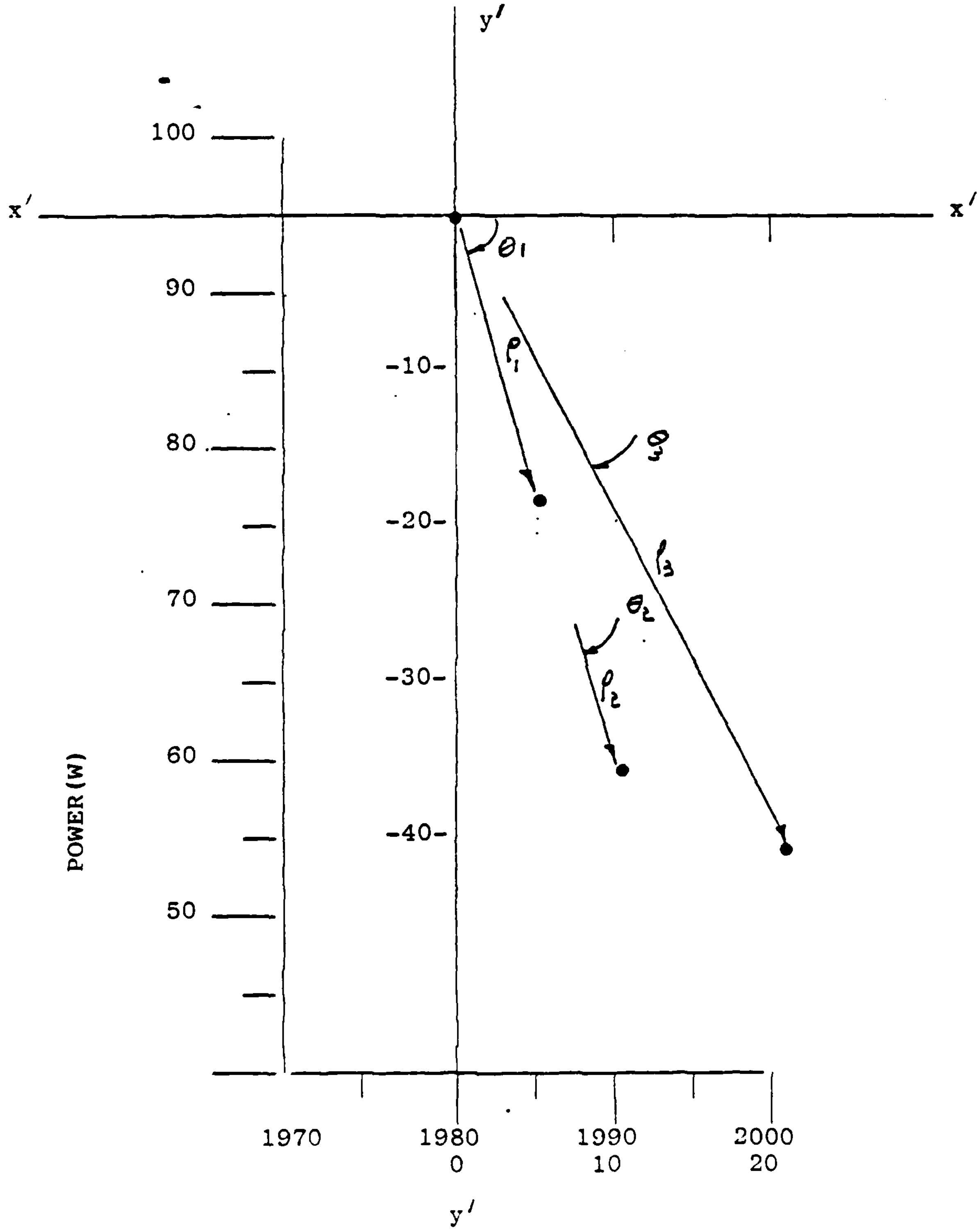


Figure 6.9

ACUMENICS

approximation and polar translation. The initial steps in altering the balanced growth forecast include:

1. establish the base year of the forecast as time zero;
2. convert the calendar years to net time beyond the base year;
3. translate the x axis from y=o to y=a;
4. calculate the slope of the likely straight line between the origin and point estimate; and
5. calculate the polar coordinates of the end point of the line.

As an example, the balanced growth scenario values for the power characteristics of VLSI instrumentation system will be translated to polar form. The original power characteristics estimate and time estimate are shown in columns a and b respectively. The translated coordinates in columns c and d are the result of altering the x-axis from y=0 y=95. The shift in the position of the x-axis does not affect the time coordinates. The calculation of the polar coordinates is accomplished using the following equations:

$$x^1 = \rho \cos \theta$$

$$y^1 = \rho \sin \theta$$

$$\rho = \sqrt{(x^1)^2 + (y^1)^2}$$

$$\theta = \arctan \frac{y}{x}$$

The steps used in the calculation of the polar point P = 18.68 θ = 164.48 are delineated below:

y^1	x^1
0	0
-18	5

ACUMENICS

The slope of the line defined by points (0,0) and (5,-18) can be calculated using

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

The slope m is equivalent to tan θ. Therefore $\theta = \arctan \frac{y_2 - y_1}{x_2 - x.}$

or in the case of interest $\theta = \arctan \frac{(-18) - (0)}{(5) - (0)} = \arctan -3.60.$

The arctan of -3.60 is 164.48°. The polar radius (ρ) is calculated using (0,0) as the origin.

Therefore $\rho = \sqrt{x^2 + y^2} = \sqrt{(5-0)^2 + (-18-0)^2} = 18.68.$

The preceeding section outlines the method for translating the unconstrained forecast into polar form. This section will discuss the rationale and technique for using the translated data to obtain modified estimates of technology characteristics and time. The assumptions adopted in Sections 6.3 and 6.4 indicate that:

a) technological change is evolutionary;

b) technological change is related to productivity; and,

c) a surrogate measure for productivity is the technology acceleration factors (A).

Additional assumptions to be used in the modification of forecast estimates include:

a) a departure from the balanced growth scenario will be measured as a change in the polar radius slope of the technology characteristics;

b) the relationship between the balanced growth slope (M) and the rapid growth (N) or stagflation scenario (P) is given by:

$$N = \frac{M}{C_N} \text{ or } P \quad \frac{M}{C_P}$$

where C_N, C_P, are defined in Figure 6.4.

The following sample calculations use the data from Figure 6.10 as well as the translation data presented above. The purpose of the example is to show the techniques for altering the balanced growth forecast to conform to either the stagflation or rapid growth scenarios.

BALANCED GROWTH

TIME	POWER	M	ρ	θ
0	95			
5	77	-3.60	18.68	164.48

The sample characteristic will convert the balanced growth five year power estimate to the stagflation estimate. The value of C_P is 2.14 as shown in Figure 6.4.

Polar Coordinates

(a)	(b)	(c)	(d)	(e)	(f)
Original Estimates		Translated Coordinates		Polar Coordinates	
Power	Time				
Y	X	Y^1	X^1	P	θ
95	0	0	0		
77	5	-18	5	18.68	164.48
60	10	-25	10	26.93	158.20
55	20	-45	20	49.24	156.04

Figure 6.10

$$P = \frac{M}{C_P}$$

$$P = \frac{-3.60}{2.14} = -1.68$$

$$\theta^1 = \arctan -1.68$$

$$\theta^1 = 149.24°$$

The value of the polar angle for the stagflation scenario is 149.24°. The polar radius remains 18.68 since it is independent of the polar angle. The remaining steps in the calculation of time and technology parameters for the data include:

a) the conversion of the polar data to the intermate translated coordinates; and,

b) conversion of the translated coordinates to the original cartesian regime.

The conversion of the polar coordinates to transformed cartesian coordinates is as follows:

$$x^1 = \rho \cos \theta^1 = 18.68 \cos 149.24$$
$$x^1 = 9.5$$
$$y^1 = \rho\ 5 \text{ in } \theta^1 = 18.68 \sin 149.24 = -16$$

The change from transformed coordinates to the original cartesian system uses the base year data i.e. power 95; time 0. The values for power and time under the stagflation scenario are obtained by summing the results and base year data.

POWER: 95-16 = 79

TIME: 0 + 9.5 = 95

The results for the stagflation scenario are presented below in tabular form for:

YEAR	POWER
1980	95
1989.5	79

A full set of parameters for the VLSI and attendent graphics is presented inthe next section. It should be noted that this technique will alter both the time of introduction, value of a parameter or both. Decimal estimates of years are provided for convenience rather than accuracy.

6.7 Microcomputers

The unconstrained forecast for microcomputer systems is presented in Chapter 2 of this volume. Data are provided concerning the likely future value of key technological parameters for two systems: 1) a typical airborne VLSI instrumentation system and 2) a typical VLSI ground-based data processing system. The data include estimates of factors including: power (watts); volume (cubic inches), cost (constant dollars), and/or speed (MIPS). The estimates for the unconstrained forecasts of the airborne and ground based data processing system are shown in Figures 6.11 and 6.12, respectively. The unconstrained forecasts are based upon interviews with microcomputer industry experts.

VLSI INSTRUMENTATION SYSTEM CHARACTERISTICS

YEAR	CHARACTERISTICS		
	POWER (W)	VOLUME IN^3	COST (1975)
1980	95	7,910	1,444
1985	77	805	901
1990	60	675	746
2000	55	-590	585

Figure 6.11

VLSI DATA PROCESSING SYSTEM CHARACTERISTICS

YEAR	CHARACTERISTICS		
	POWER (W)	SPEED (MPS)	COST (1975)
1980	520	3.13	28,365
1985	395	7.81	7,283
1990	270	15.63	3,097
2000	145	41.67	974

Figure 6.12

The typical VLSI instrumentation system envisioned includes a display memory and controller, a 64 K-bit program memory (ROM), a 4 K-bit data memory (RAM), and a 4 K-bit bulk memory. The unconstrained forecast indicates that by 2,000 A.D. power equivalents will be reduced 42% over 1980 levels. In addition, volume requirements will be decreased by 69%. Further, unit costs will decrease by 59%.

As noted in previous sections, the unconstrained technology forecast has been deemed consistent with the balanced growth scenario. The technique described in Section 6.6 was used to develop estimates of the relevant VLSI instrumentation system technological parameters for the stagflation and rapid growth scenarios. The forecast results for the three socio-economic scenarios are shown in tabular and graphic form in Figures 6.13 to 6.18.

Power consumption, shown in Figures 6.13 and 6.14, is estimated to decline for the balanced growth scenario from 95 watts in 1980 to 55 watts in 2000, i.e a reduction of 42%. Power consumption requirements will diminish also under the stagflation scenario. However, the reduction in power requirements will be less than the 42% estimated for the balanced growth scenario. It is estimated that under the stagflation scenario power requirements will be 56.6 watts, in 2003, a reduction of 40% over 1980 levels. Power requirements under the rapid growth scenario will diminish more rapidly than under balanced growth. It is estimated that during

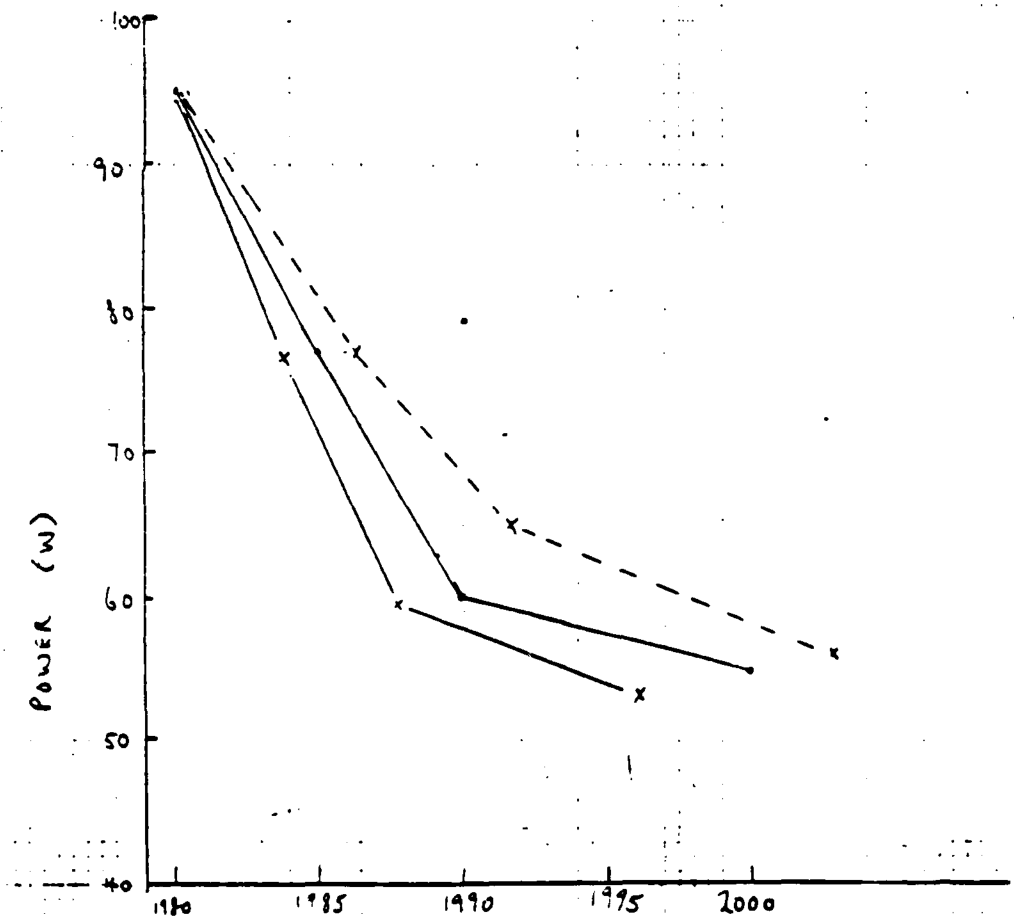
VLSI INSTRUMENTATION SYSTEM

POWER

BALANCED GROWTH		STAGFLATION		RAPID GROWTH	
YEAR	POWER (w)	YEAR	POWER	YEAR	POWER
1980	95	1980	95	1980	95
1985	77	1985.9	77.3	1983.9	76.7
1990	60	1991.8	60.6	1987.8	59.5
2000	55	2003	56.6	1996.1	53.3

Figure 6.13

VLSI INSTRUMENTATION SYSTEM
POWER REQUIREMENT FORECAST



KEY

●———● Balanced Growth

x------x Stagflation

x———x Rapid Growth

Figure 6.14

ACUMENICS

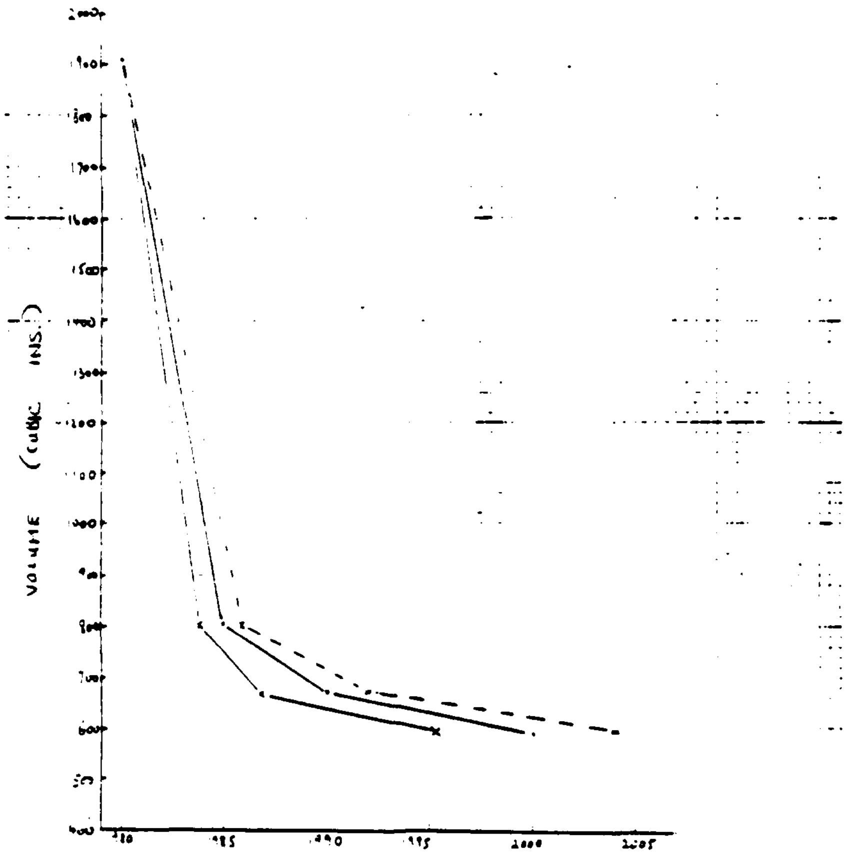
VLSI INSTRUMENTATION SYSTEM

Volume

BALANCED GROWTH		STAGFLATION		RAPID GROWTH	
YEAR	VOLUME (cu. in.)	YEAR	VOLUME (cu. in.)	YEAR	VOLUME (cu. in.)
1980	1910	1980	1910	1980	1910
1985	805	1986	805	1984	805
1990	675	1992	675	1987	675
2000	590	2004	590	1995.4	590

Figure 6.15

VLSI INSTRUMENTATION SYSTEM
VOLUME FORECAST



KEY

o———o Balanced Growth

x------x Stagflation

x———x Rapid Growth

Figure 6.16

ACUMENICS

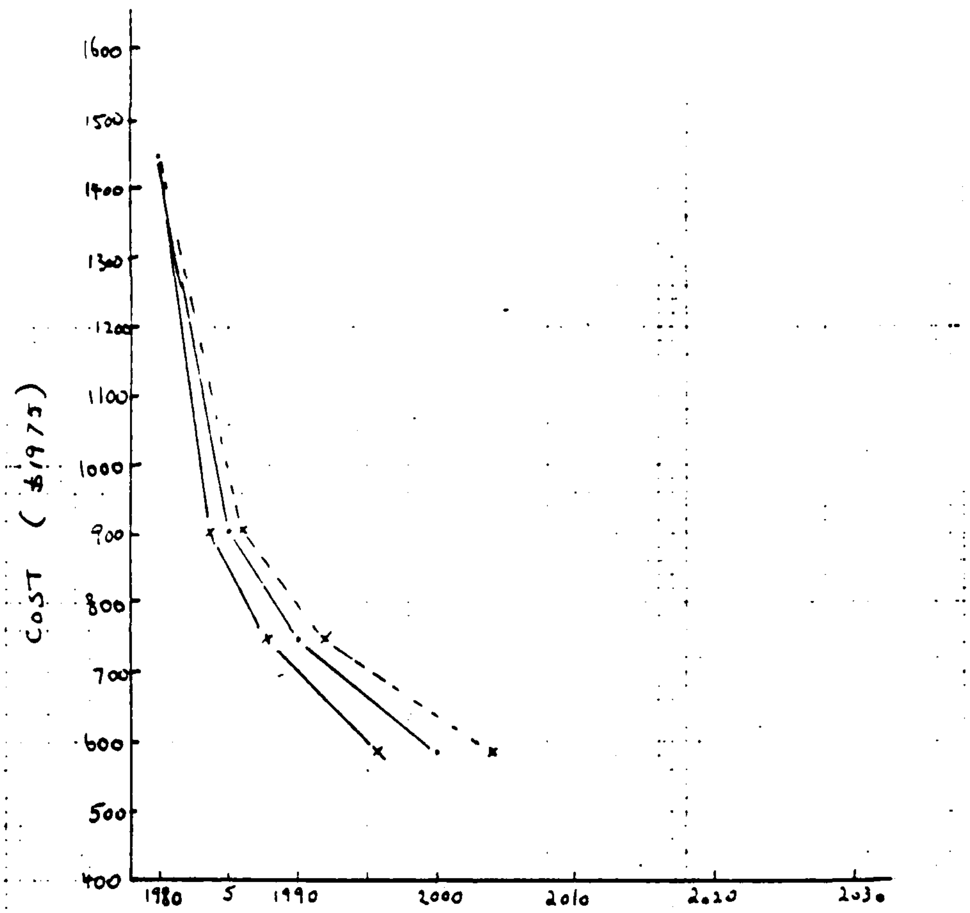
VLSI INSTRUMENTATION SYSTEM

COST
(1975 $)

BALANCED GROWTH		STAGFLATION		RAPID GROWTH	
YEAR	COST	YEAR	COST	YEAR	COST
1980	1444	1980	1444	1980	1444
1985	901	1986	901	1983.9	901
1990	746	1992	746	1987.7	746
2000	585	2004	585.1	1995.4	584.9

Figure 6.17

VLSI INSRUMENTATION SYSTEM
UNIT COST FORECAST



K_Y

o———o Balanced Growth

x------x Stagflation

x———x Rapid Growth

Figure 6.18

ACUMENICS

1996 under the rapid growth scenario power requirements will be 53.3 watts, or 44% less than 1980 levels.

The volume required for the VLSI instrumentation system will diminish under all three socio-economic scenarios, but at different rates. Under the balanced growth scenario space requirements will decrease from 1910 cubic inches to 590 cubic inches in 2000. The 590 cubic inches represents a 69% reduction. The same reduction, i.e. 69%, is anticipated under the rapid growth and stagflation scenarios. However, 590 cubic inch size will not be achieved until 2004 and 1995, under the stagflation and rapid growth scenarios, respectively.

VLSI system cost will decrease in each of the three scenarios. The balanced growth scenario will result in system costs decreasing from $1,444 in 1980 to 585 in 2,000. The preceeding represents a reduction of 59%. The absolute reduction in system cost in 2000 under the balanced growth scenario will not be attained until 2004 if the stagflation construct is in force. The system cost under the stagflation scenario during 2000 is estimated at 639, a 55.5% reduction from 1980 costs. It is expected that the balanced growth scenario systems costs for the year 2000, $585, will be attained four to five years sooner under the rapid growth scenario.

VLSI Land Based Data Processing System

The forecast data included in this section are based upon a typical VLSI data processing system equivalent to an Amdahl 470 or IBM 370 unit. The prototype VLSI data processing system will have 200,000 gates in the CPU and 32 M - bits of read-write high speed memory. The unconstrained technology forecast includes estimates of power (watts), speed (MIPS) and cost. The forecast estimates for the VLSI data processing system characteristics are presented in tabular form in Figure 6.19, 6.21, 6.23 and graphical form in Figures 6.20, 6.22, and 6.24.

The power requirements for the VLSI data processing system will follow the trends delineated for the VLSI instrumentation technology; i.e., they will be diminished over the forecast period. In the approximate 20 year period, a 72% reduction will be realized, declining from 520 watts in 1980 to 145 watts in 2000. Power requirements under the stagflation scenario will not achieve the same level of reduction as the balanced growth scenario, until 2004. It is estimated that the power requirements for the VLSI data processing system under stagflation during 2000 will be 187 watts, equivalent to a 64% reduction over the 1980 level. The 145 watt power requirement level for the rapid growth scenario will be attained four to five years earlier than projected for the balanced growth scenario.

The ability to process data or the speed of the VLSI data processing system is likely to increase rapidly under all three

VLSI DATA PROCESSING SYSTEM CHARACTERISTICS

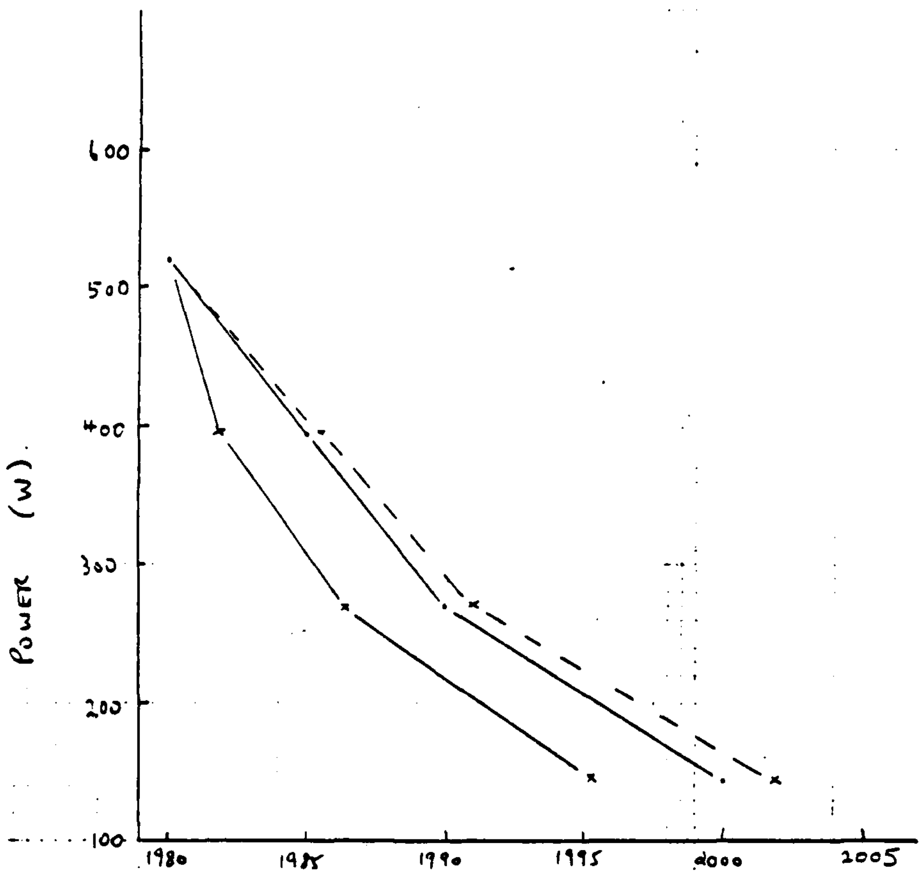
POWER (W)

SCENARIO					
BALANCED GROWTH		STAGFLATION		RAPID GROWTH	
YEAR	SIZE (Watts)	YEAR	SIZE (Watts)	YEAR	SIZE (Watts.)
1980	520	1980	520	1980	520
1985	395	1986	394.9	1983.8	395
1990	270	1992	270.1	1987.7	270.9
2000	145	2004	145.2	1995.4	144.8

Figure 6.19

ACUMENICS

VLSI DATA PROCESSING SYSTEM CHARACTERISTICS
POWER REQUIREMENTS FORECAST



KEY

o———o Balanced Growth

x-------x Stagflation

x———x Rapid Growth

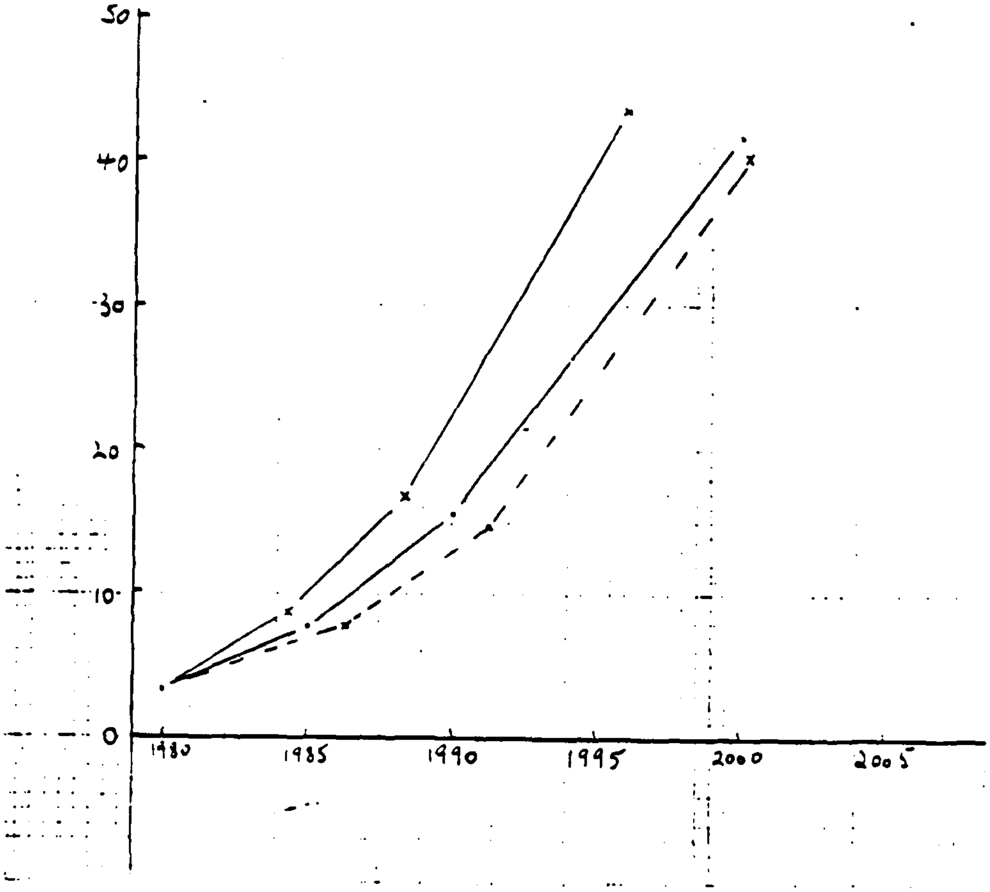
Figure 6.20

SPEED (MIPS)

SCENARIO					
BALANCED GROWTH		STAGFLATION		RAPID GROWTH	
YEAR	S	YEAR	S	YEAR	S
1980	3.1	1980	3.1	1980	3.1
1985	7.8	1985.4	7.3	1984.3	8.4
1990	15.6	1991.1	14.7	1988.4	16.8
2000	41.7	2002.9	40	1996.1	43.5

Figure 6.21

VLSI DATA PROCESSING SYSTEM
SPEED FORECAST



KEY

o———o Balanced Growth

x------x Stagflation

x———x Rapid Growth

Figure 6.22

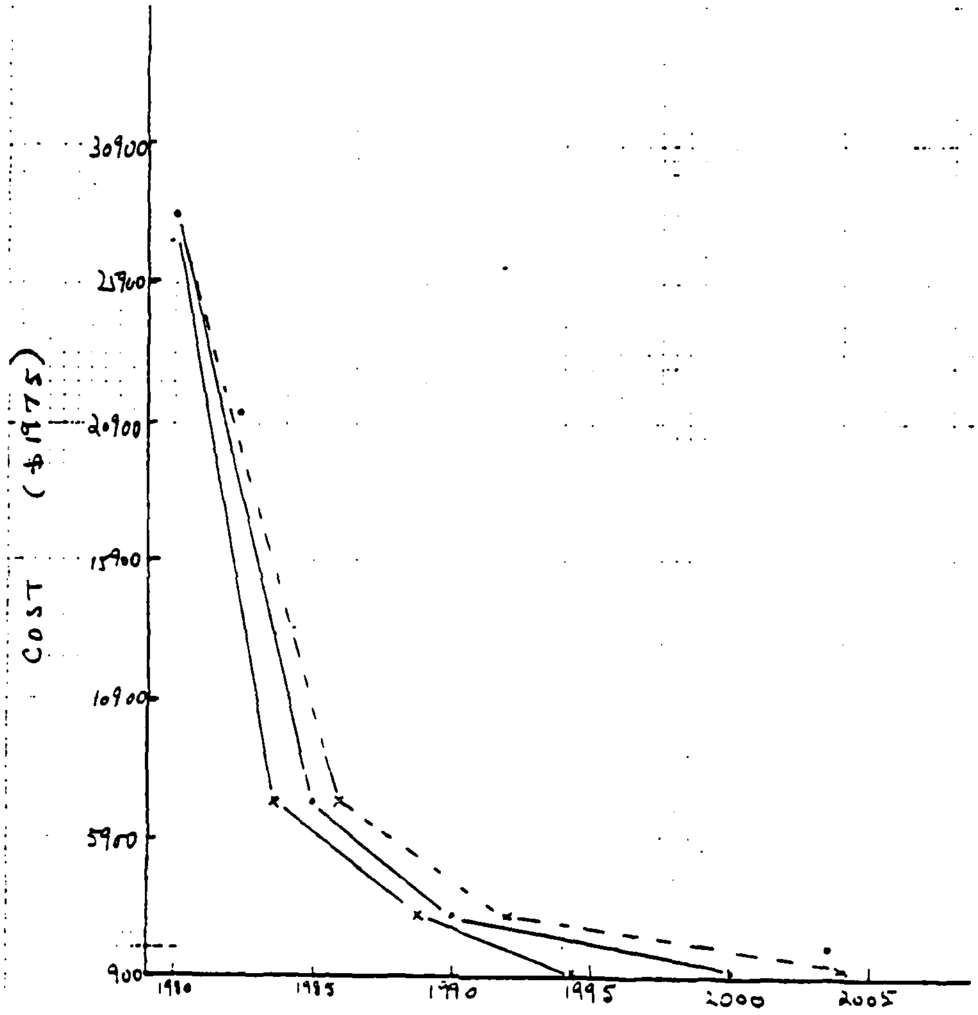
VLSI DATA PROCESSING

COST
(1975 $)

SCENARIO					
BALANCED GROWTH		STAGFLATION		RAPID GROWTH	
YEAR	COST	YEAR	COST	YEAR	COST
1980	28,365	1980	28,365	1980	28,365
1985	7,283	1986	7,283	1983.7	7,283
1990	3,097	1992	3,097	1988.8	3,097
2000	974	2004	974	1994.3	974

Figure 6.23

VLSI DATA PROCESSING SYSTEM
COST FORECAST



KEY

o———o Balanced Growth

x------x Stagflation

x———x Rapid Growth

Figure 6.24

ACUMENICS

scenarios. The unconstrained or balanced growth forecast is estimated to increase from 3.1 to 41.7 MIPS between 1980 and 2000, a change of 1245%. Increased data processing speed will occur also in the stagflation scenario. However, the speed attained during 2000 under stagflation will be 34.2 MIPS, an increase of 1103%. The greatest increase in system speed will occur under the rapid growth scenario. It is anticipated that by 1996.1 system speed will attain 43.5 MIPS under the rapid growth scenario.

The cost of the VLSI data processing system will decrease most rapidly. A decrease of 97%, from $28,365 in 1980 to $974 in 1994.3 is forecast for the rapid growth scenario. Under the balanced growth scenario the 97% decline in VLSI data processing cost will occur by 2000. While the same percentage will not be realized until 2004 for the stagflation growth scenario.

6.8 Input-Output Devices

The typical prototype input-output device considered in the technology forecast is comprised of a microcomputer control complex with a 1 o M-bit RAM, a 75 M-byte tape bulk storage module, a hardcopy output device, a TV-quality video input system, a 25 inch video display screen, and standard audio and keyboard I/O devices.

It is expected that the technology of input-output devices will undergo great change independent of the socio-economic scenarios. The scenarios, however, will impact the rate of such change.

The impact of the scenarios are estimated in terms of size (cubic inches) and cost.

The estimates of the forecasted technology parameters for each scenario are displayed in tabular form in Figures 6.25, 6.27 and in graphic form in Figures 6.26 and 6.28.

The input-output device is expected to decrease 65% from the 1980 size of 39,700 cubic inches to 13,740 cubic inches. Under the rapid growth scenario this decrease is forecast to occur in 1995.4, some four to five years earlier than that of the balanced growth scenario which is expected to decline to the same size by 2000. Under the stag[illegible]tion scenario the input-output device is not expected to attain 13,740 cubic inches until the latter half of 2002.

The cost of input-output devices are expected to follow the patterns established for unit size. That is, an 87% reduction is expected to occur, declining from $19,626 in 1980 to $2,491 by 1995, 2000 and 2004 for the rapid growth, balanced growth and stagflation growth scenarios, respectively.

6.9 Switching Systems

The nature and unconstrained forecast of switching systems is contained in Chapter 4. It should be noted that the discussion in Chapter 4 concentrated on telephone network switching systems.

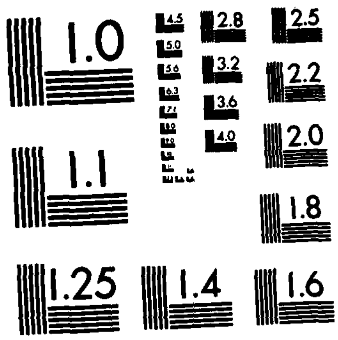
INPUT-OUTPUT TERMINAL CHARACTERISTICS

SIZE (in^3)					
BALANCED GROWTH		STAGFLATION		RAPID GROWTH	
YEAR	SIZE	YEAR	SIZE	YEAR	SIZE
1980	39,700	1980	39,700	1980	39,700
1985	33,550	1986	33,550	1983.9	33,550
1990	28,000	1992	28,000	1987.7	28,000
1995	14,450	1997.6	14,450	1991.6	14,450
2000	13,740	2002.7	13,740	1995.4	13,740

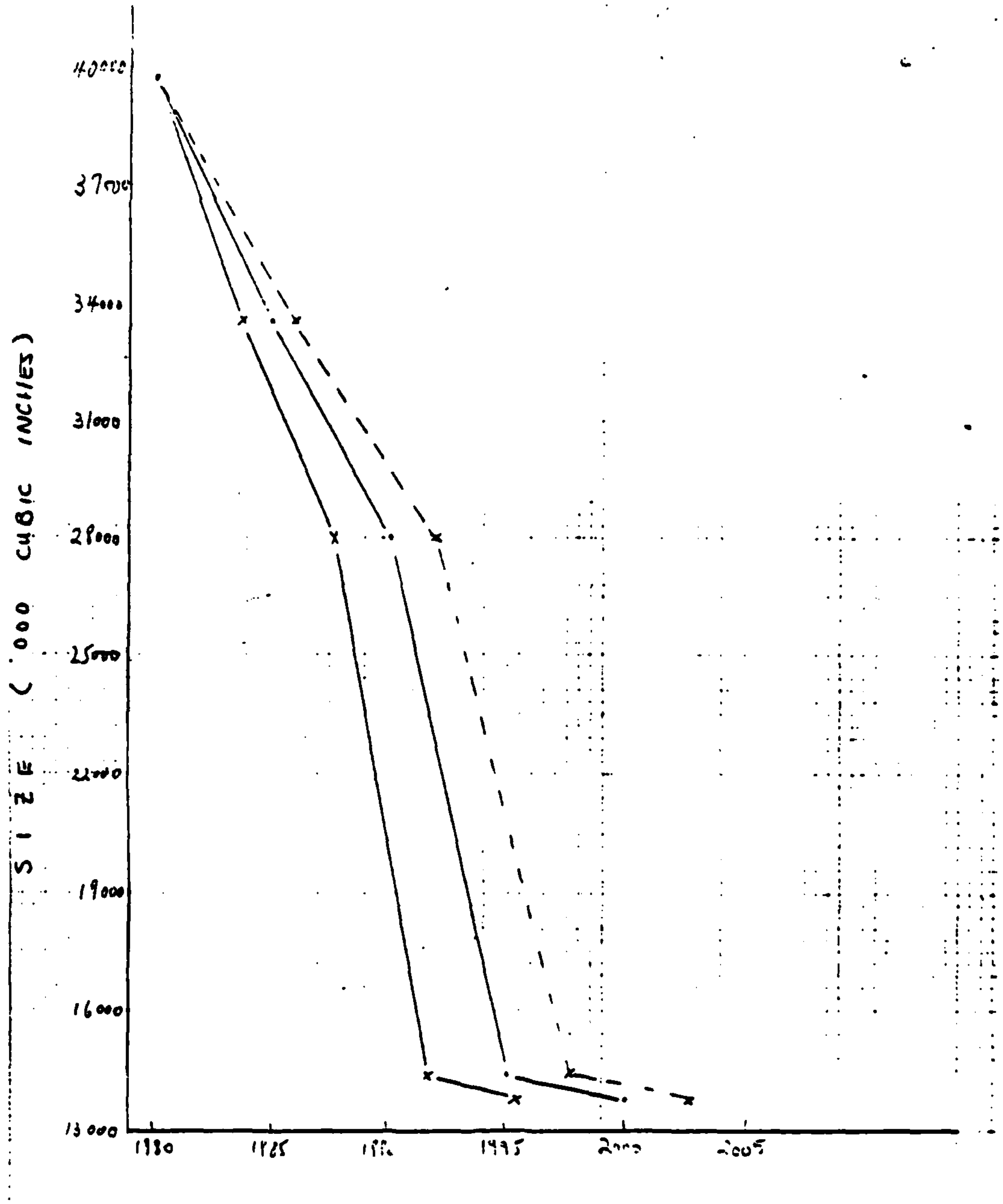
Figure 6.25

AD-A126 015 SOCIOECONOMIC IMPACT ASSESSMENT: COMMUNICATIONS INNUSTRY PHASE III TECHNO..(U) ACUMENICS RESEARCH AND TECHNOLOGY INC BETHESDA MD 02 FEB 79 FAA-APO-81-11-4 6/7

UNCLASSIFIED DOT-FA78WAI-932 F/G 17/2.1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



KEY

o———o Balanced Growth

x------x Stagflation

x———x Rapid Growth

Figure 6.26

ACUMENICS

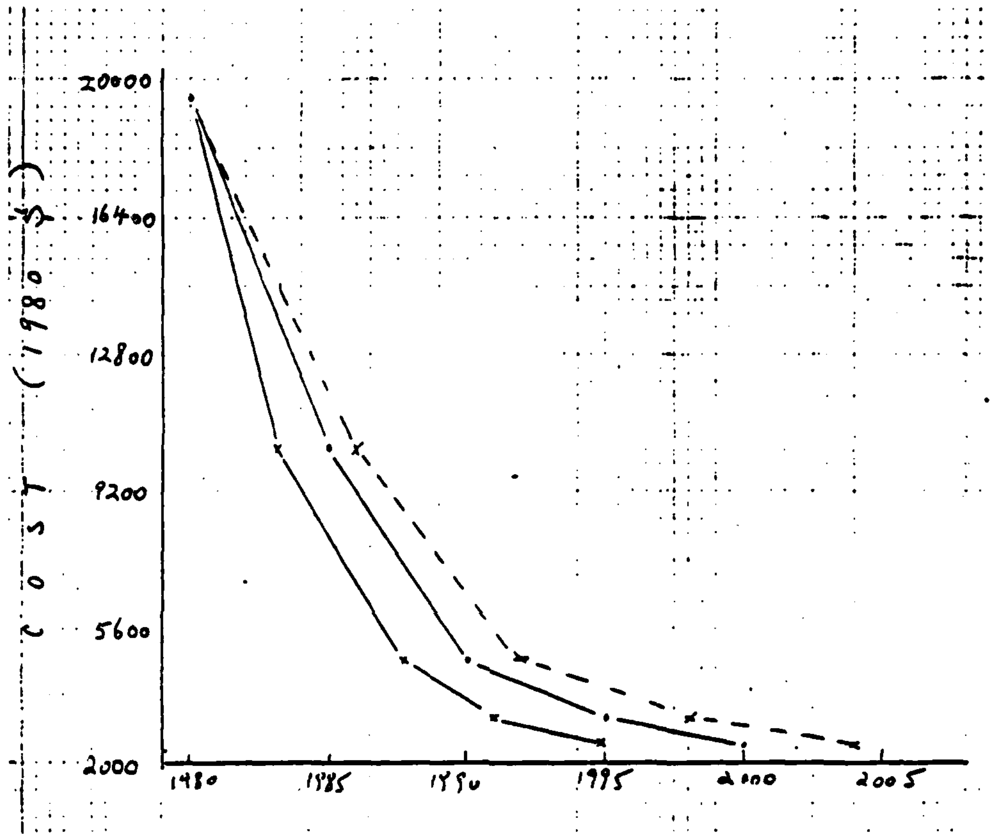
INPUT-OUTPUT TERMINAL CHARACTERISTICS

COST
(1980 $)

BALANCED GROWTH		STAGFLATION		RAPID GROWTH	
YEAR	COST	YEAR	COST	YEAR	COST
1980	19,626	1980	19,626	1980	19,626
1985	10,314	1986	10,314	1983.3	10,314
1990	4,835	1992	4,835	1987.7	4,835
1995	3,146	1998	3,146	1991.5	3,146
2000	2,491	2004	2,491	1995	2,491

Figure 6.27

INPUT-OUTPUT TERMINAL COST FORECAST



KEY

o———o Balanced Growth

x------x Stagflation

x———x Rapid Growth

Figure 6.28

However, pocket switching and other enhanced services were briefly discussed. Further, the forecast contained in Chapter 4 was qualitative rather than quantitative. That is, estimates of the mix of switching systems were provided rather than numeric projections of technology parameters. As such, it is necessary to cast this section in a manner different from the previous section of Chapter 6. The principle difference will be in the nature of the parameter rather than the technique of forecast.

The data obtained is the unconstrained i.e. balanced growth forecast as shown in Figure 6.31. The data provided are a forecast of the proportion of different switching systems in service for specific years. It was assumed that the parameter of impact is the proportion of electronic switching systems in service. The proportion of electronic switching systems in service was forecasted for the rapid growth and stagflation scenarios with the technique delineated in Section 6.4.

The forecasted proportion of electronic switching systems for all scenarios is shown in Figures 6.29 and 6.30. The balanced growth scenarios indicated that by 1995 all switching will be electronic. However, it is estimated that under the stagflation scenario complete electronic switching will not occur until after 1997. It is anticipated that under the rapid growth scenario electronic switching will be in full force during the latter months of 1991.

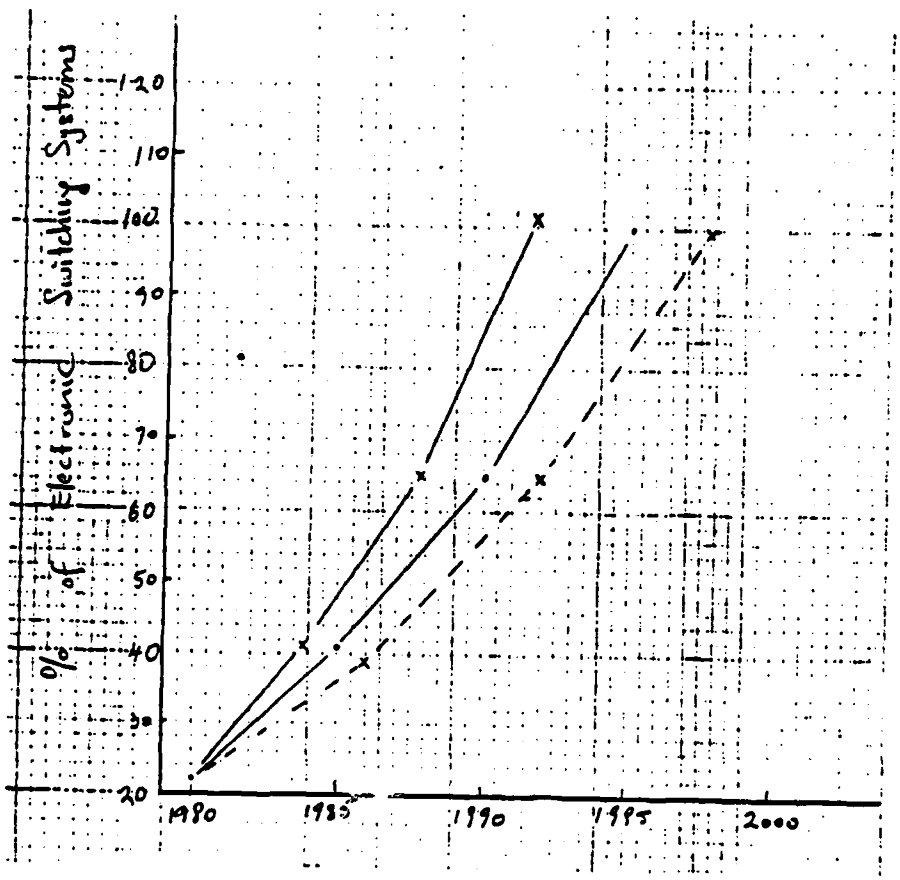
SWITCHING SYSTEMS

FORECAST OF ELECTRONIC PERCENTAGE

BALANCED GROWTH		STAGFLATION		RAPID GROWTH	
YEAR	PERCENTAGE	YEAR	PERCENTAGE	YEAR	PERCENTAGE
1980	22	1980	22	1980	22
1985	41	1985.9	38.7	1983.9	41.3
1990	65	1991.9	64.5	1987.8	65.5
1995	100	1997.9	99.4	1991.7	100.6

Figure 6.29

SWITCHING SYSTEMS
FORECAST OF PERCENTAGE ELECTRONIC



KEY

o———o	Balanced Growth
x------x	Stagflation
x———x	Rapid Growth

Figure 6.30

ACUMENICS

MIX OF SWITCHING SYSTEMS

BALANCED GROWTH

Year	Electronic	#5 Cross Bar	#1 Cross Bar	Panel	Step By Step
1980	22	43	7	2	26
1985	41	33	5	0	21
1990	65	20	3	0	12
1995	100	0	0	0	0

Figure 6.31

The relative mix of switching systems for each scenario are shown in Figures 6.31, 6.32 and 6.33. The relative mixes are based on subjective allocation, based on the forecast of electronic switching in Figure 6.29.

6.10 Transmission System

Unconstrained forecasts of future communication transmission systems are presented in Chapter 5. The most important transmission systems are based upon fiber optic and satellite technology. The unconstrained forecast of fiber optic technology is used primarily upon technical considerations and limitations. The unconstrained forecast for satellite technology derives from mathematical extrapolation with respect to the features of current technology.

Fiber optics remains a fledgling technology. As such, parameters for characterizing the technology are yet to be developed. For the purpose of this forecast, estimates have been provided for the cost of fiber optic communication units. The forecasted cost values for fiber optic levels are shown in tabular form and graphic form, Figures 6.34 and 6.35 respectively. The unconstrained forecast values are shown in the balanced growth section of Figure 6.34. If the balanced growth scenario obtains, then fiber optic link costs are expected to decline from $11,000 in 1980 to $570 in 2000, 95% by 1993.1. An equivalent comparison between the balanced growth rapid growth extrapolation indicates that a 95% decline in cost will be achieved during 1998 under the rapid growth scenario, almost two years earlier than forecast for the balanced growth scenario.

MIX OF SWITCHING SYSTEMS

STAGFLATION

Year	Electronic	#5 Cross Bar	#1 Cross Bar	Panel	Step By Step
1980	22	43	7	2	26
1989.6	39	35	5	0	21
1999.7	62	33	3	0	12
2010.0	95	5	0	0	0

Figure 6.32

MIX OF SWITCHING SYSTEMS

RAPID GROWTH

Year	Electronic	#5 Cross Bar	#1 Cross Bar	Panel	Step By Step
1980	22	43	7	2	26
1983.3	41	33	5	0	21
1986.6	66	19	3	0	12
1989.9	100	0	0	0	0

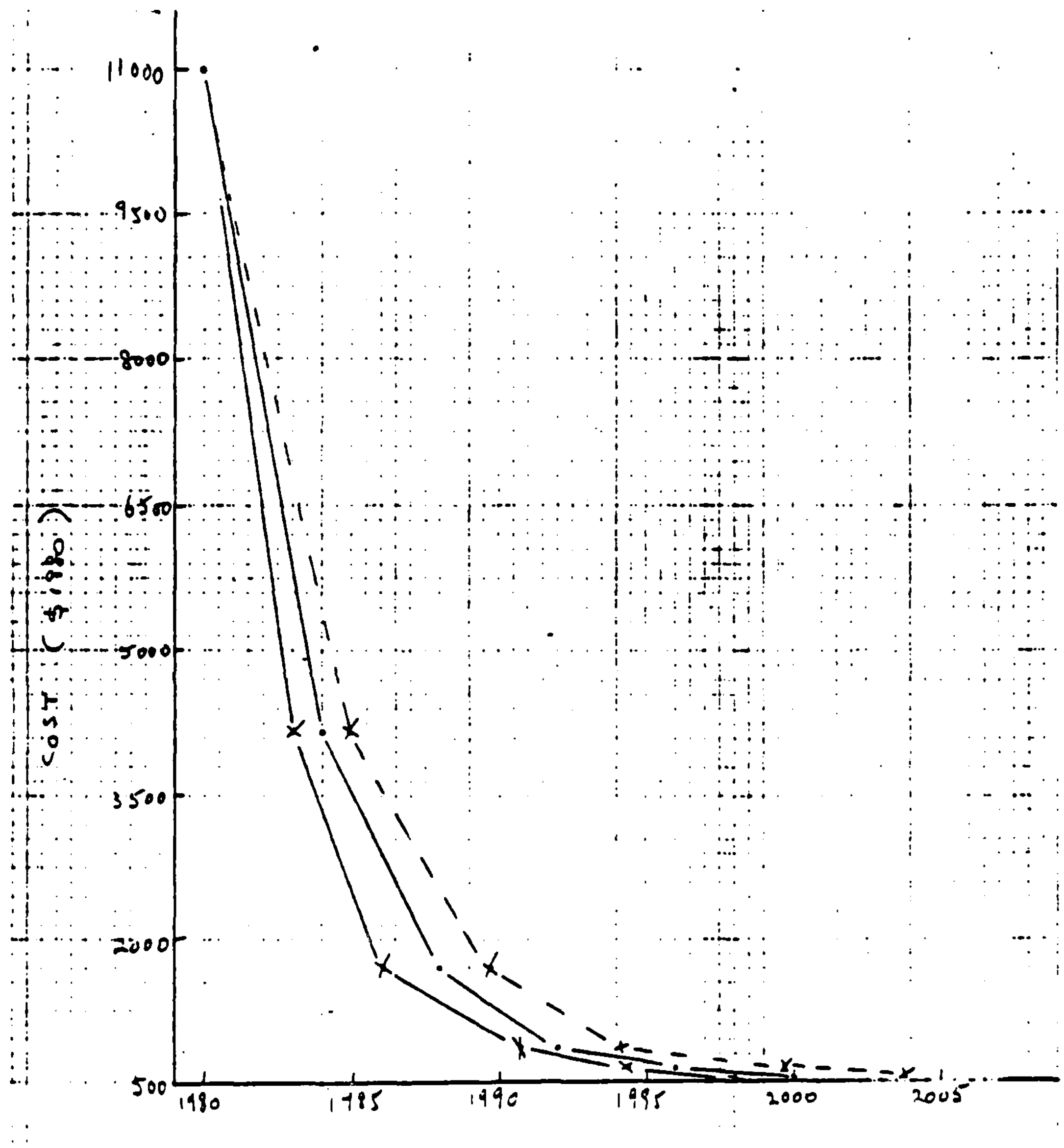
Figure 6.33

FIBER OPTIC LINK COST (1980 $)

BALANCED GROWTH		STAGFLATION		RAPID GROWTH	
YEAR	COST($)	YEAR	COST($)	YEAR	COST($)
1980	11,000	1980	11,000	1980	11,000
1984	4,150	1984.8	4,150	1983.6	4,150
1988	1,700	1989.7	1,700	1986.5	1,700
1992	880	1994.1	880	1990.6	880
1996	650	1999.9	650	1994.5	650
2000	570	2003.7	570	1998.2	570

Figure 6.34

FIBER OPTIC LINK
COST FORECAST



KEY

o———o Balanced Growth

x------x Stagflation

x———x Rapid Growth

Figure 6.35

As expected the 95% decline will be achieved at the latest if the stagflation scenario should hold. The price of $570 will not be attained until sometime during the latter half of 2003 for this scenario, 5-6 years later than forecast for the rapid growth scenario, and 3-4 years after that of the balanced growth scenario.

The unconstrained forecast for satellite technology projects values for weight in orbit and number of wide band channels per unit (Figure 6.36 and 6.37). Under the unconstrained or balanced growth scenario the weight in orbit is expected to increase from 14,500 KG in 1980 to 67,583 KG in 2000, or 366%. The 366% gain in orbit weight is anticipated by 1995.4 under the rapid growth scenario; not until 2004 under the stagflation scenario.

The number of wide band channels under the balanced growth scenario is expected to increase from 612 in 1980 to 3822 in 2000. (Figures 6.38 and 6.39). It is anticipated that 2725 channels will be available per satellite by 1995 under balanced growth, by 1998 under stagflation, and by 1991.1 under rapid growth. The estimated leasing cost per channel for the three scenarios is given in Figures 6.40 and 6.41.

COMMUNICATION SATELLITE

CHARACTERISTICS

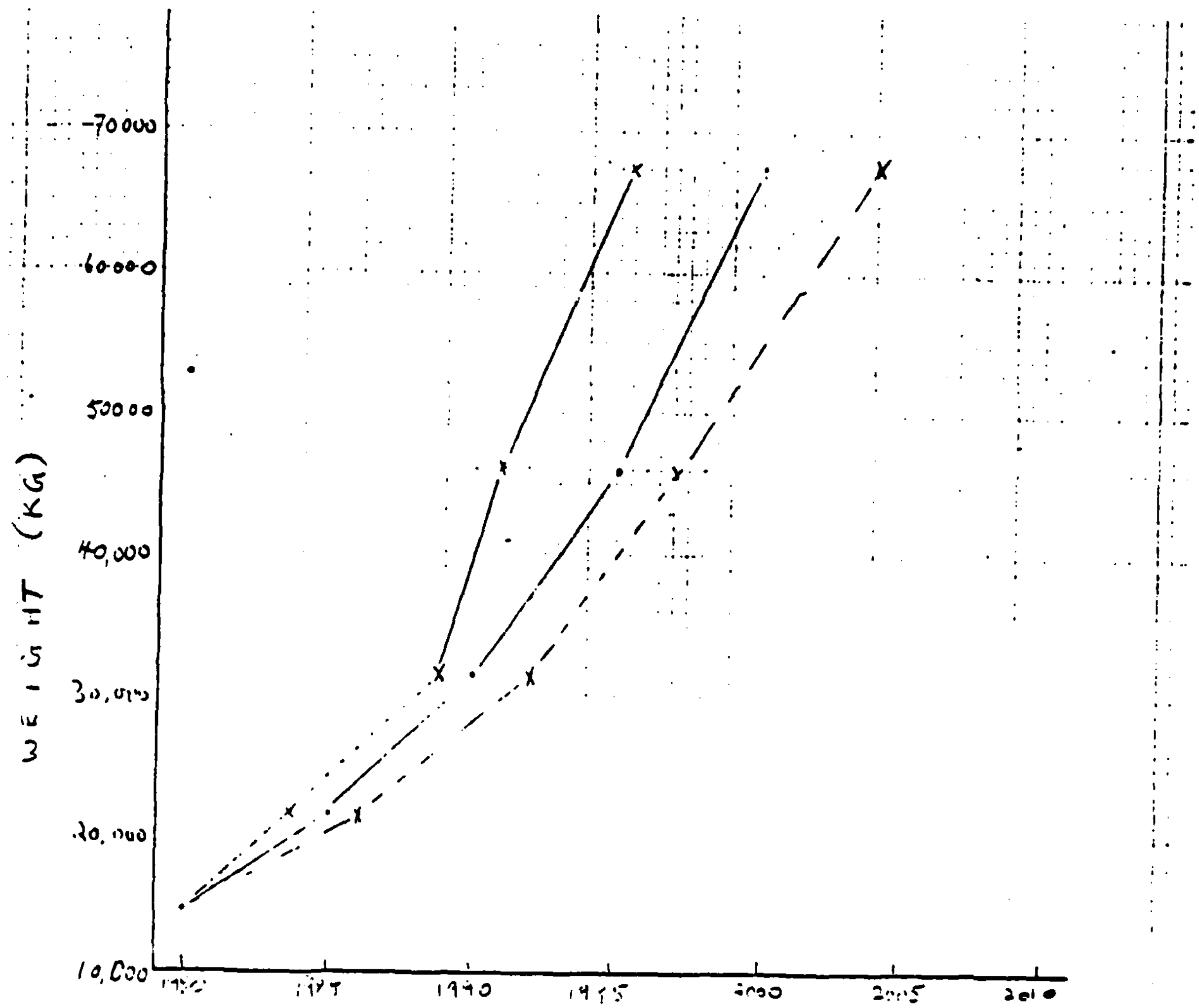
WEIGHT

(WEIGHT (KG))

BALANCED GROWTH		STAGFLATION		RAPID GROWTH	
YEAR	WEIGHT	YEAR	WEIGHT	YEAR	WEIGHT
1980	14,500	1980	14,500	1980	14,500
1985	21,305	1986	21,305	1983.6	21,305
1990	31,304	1992	31,304	1988.8	31,304
1995	45,996	1997	45,996	1991	45,996
2000	67,583	2004	67,583	1995.4	67,583

Figure 6.36

COMMUNICATION SATELLITE
WEIGHT FORECAST



KEY

o———o Balanced Growth

x------x Stagflation

x———x Rapid Growth

Figure 6.37

ACUMENICS

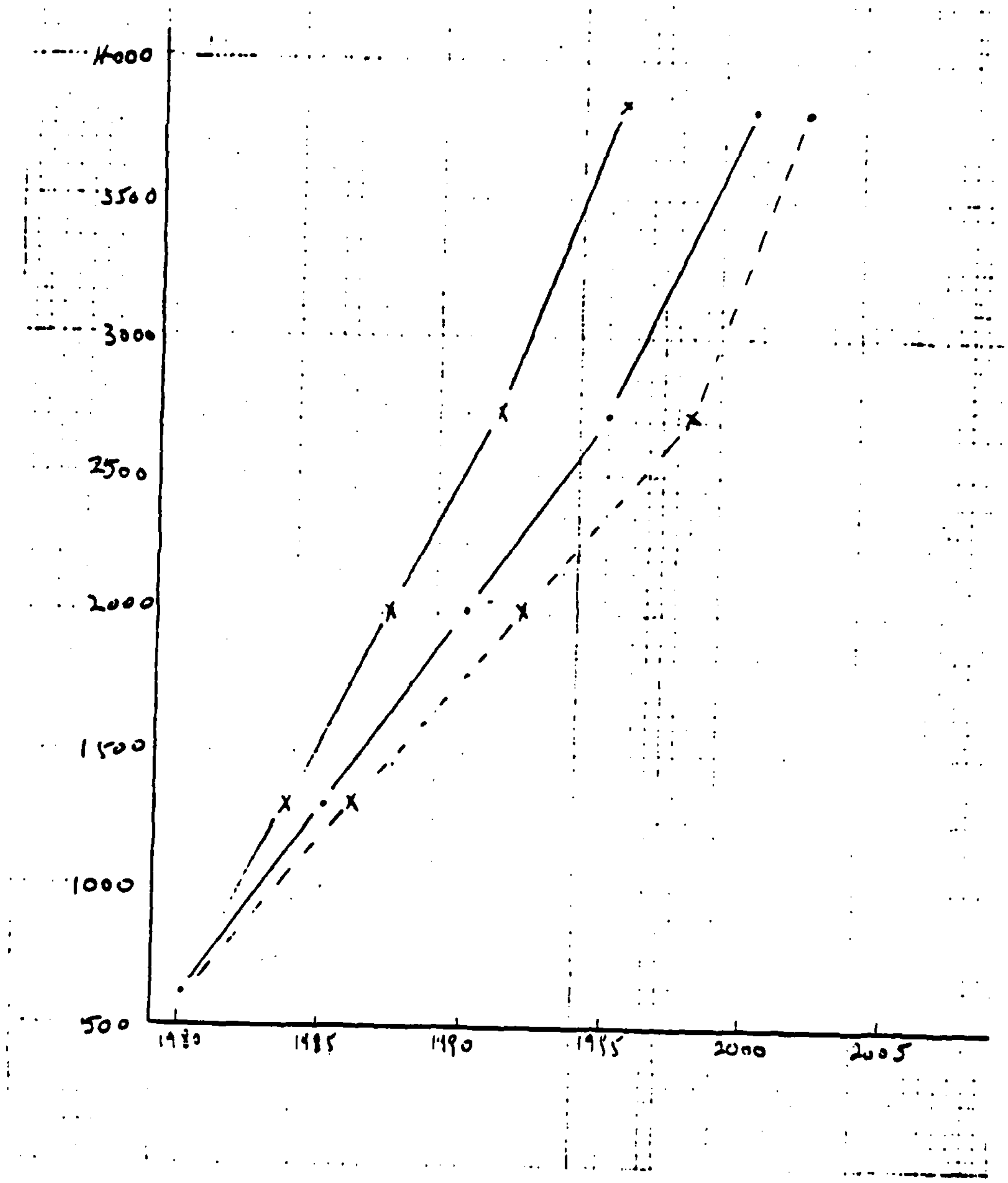
NUMBER WIDE BAND

CHANNELS PER SATELLITE

BALANCED GROWTH		STAGFLATION		RAPID GROWTH	
YEAR	CHANNEL	YEAR	CHANNEL	YEAR	CHANNEL
1980	612	1980	612	1980	612
1985	1310	1984.7	1081	1983.6	1391
1990	2005	1989.4	1548	1987.3	2166
1995	2725	1994.3	2037	1990.9	2965
2000	3822	2002.7	2881	1994.2	4117

Figure 6.38

NUMBER OF WIDE BAND CHANNELS/SATELLITES



KEY

o———o Balanced Growth

x------x Stagflation

x———x Rapid Growth

Figure 6.39

ACUMENICS

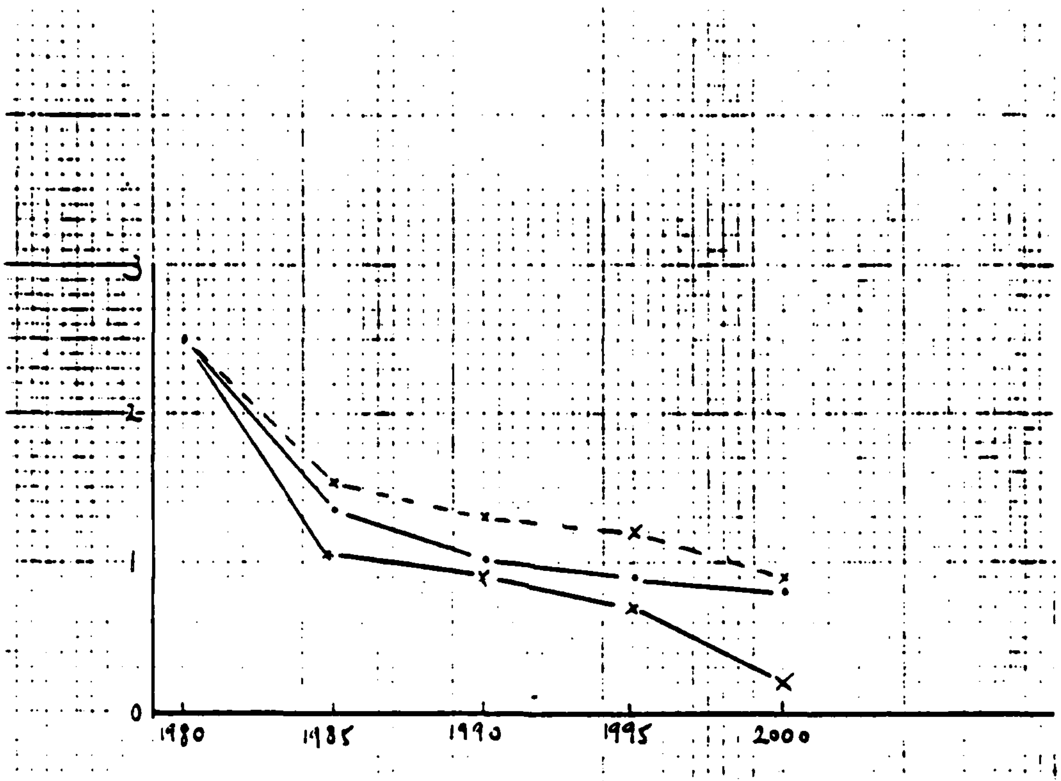
SATELLITE CHARACTERISTICS

LEASING COST/CHANNEL ($M/YEAR)

BALANCED GROWTH		STAGFLATION		RAPID GROWTH	
YEAR	COST	YEAR	COST	YEAR	COST
1980	2.5	1980	2.5	1980	2.5
1985	1.35	1985	1.54	1984.9	1.03
1990	1.02	1990	1.3	1990	.9
1995	.9	1995	1.2	1995	.7
2000	.8	2000	.9	2000	.2

Figure 6.40

SATELLITE LEASING
COST PER CHANNEL



KEY

o———o Balanced Growth

x------x Stagflation

x———x Rapid Growth

Figure 6.41

ACUMENICS

7.1 Introduction

The object of this paper is to define the elements of communication systems that may be available to serve the needs of civil aviation for the year 2020. The prospective communication systems are described in this section. The specifications for each system derive from the socio-economic scenarios developed in Phase II as well as the technology forecast prepared in Phase III. One communication system is delineated for each economic scenario. The communication system concept for the rapid growth scenario is primarily space based. The system requirement under the balanced growth scenario is a blend of terrestrial and space-based facilities. Under the stagflation scenario the communication system will be primarily terrestrially based.

The proposed systems will utilize the technology in widespread use in the year 2000, because the diffusion of communication technology requires 20 years. In addition, the economic and technological life of avionics equipment is approximately 10 years. The negotiation and adoption of national agreements for new navigation systems requires at least 5 years. The new system will require design and development which will require five years. Thus, a 20-year time lag between technological availability and widespread use may be optimistic. However, the time lag in adoption identified above to some extent does reduce uncertainty associated in this system design; that is, a forecast of technology to be available in 20 years will be more certain than such projections for 40 years ahead.

It should be noted that the communication system will be a function of not only the technology available, but also user demand. For this reason, the system configurations presented in succeeding sections may be subject to major changes resulting from fluctuations in demand. Further the system concepts assume that the functions of the National Airspace System (NAS) will remain constant through 2020. For this purpose, the functions of the NAS will be:

1. to provide aircraft with a means for navigation;
2. to monitor the position of airborne aircraft for surveillance and ground management to assure their separation;
3. to provide both voice and data communication links between aircraft and ground stations;
4. to provide ground-to-ground communication.

A change in NAS functions will also influence system design. At present the NAS functions are performed using terrestrially-based facilities. A significant portion of the demand for communication services derives from the need to provide information transfer among the terrestrially-based facilities. As an example, there are three major dedicated FAA terrestrial networks for FAA data communications: the Aeronautical Fixed Telecommunications Network (AFTN), the Weather Teletypewriter Networks Service A, C and O, and the Service B System (Figure 7.1). The three networks carry national and international digital message traffic in support of air traffic control operations and aeronautical weather services. Present agency plans indicated that the three networks will be integrated into the National Airspace Data Exchange Network (NADIN) (Figure 7.2) in the mid-nineteen-eighties.

Design for Modernized Service B Network with 1.5 times 1984 Traffic Requirements

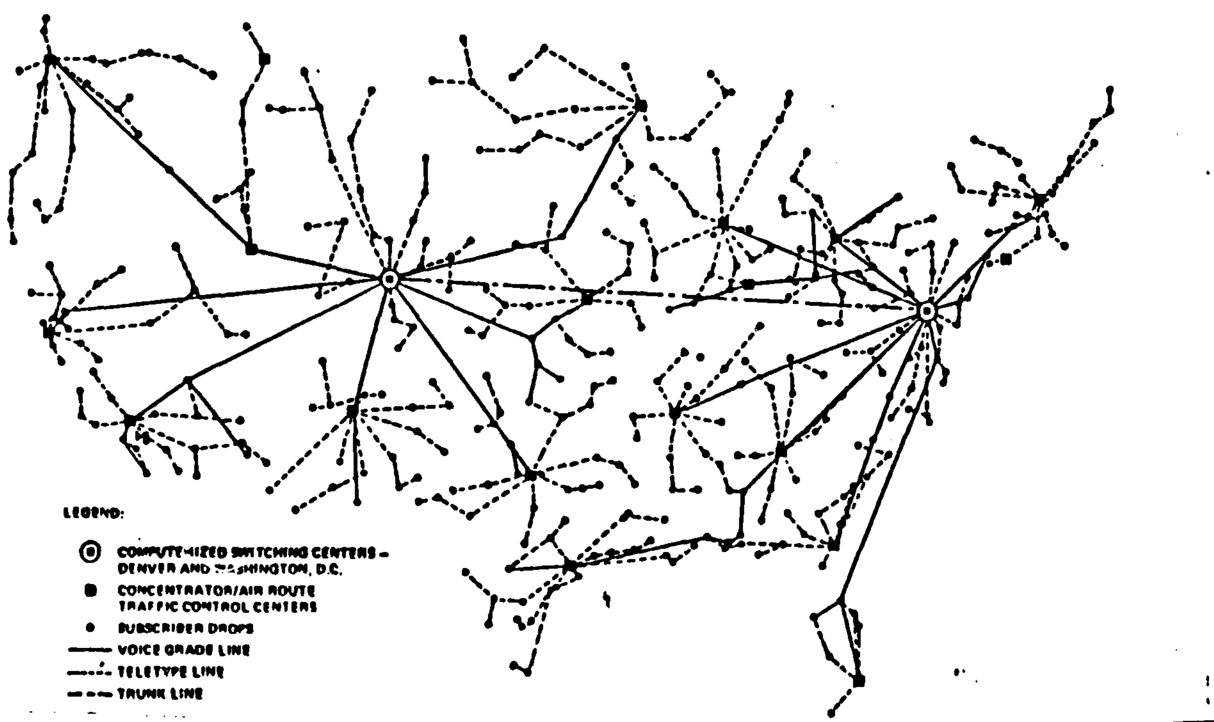
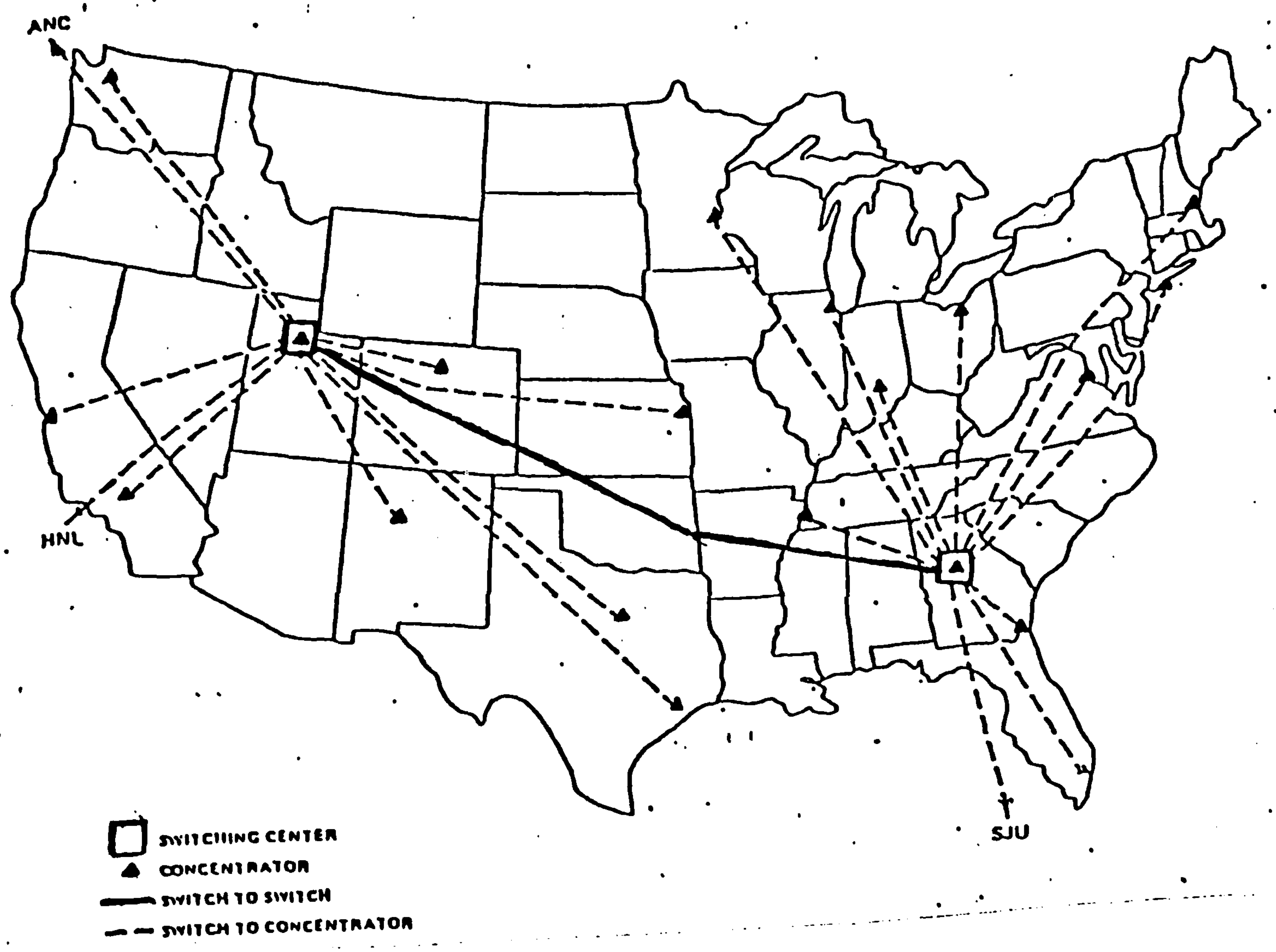


Figure 7.1

National Airspace Data Interchange Network (NADIN)

Back Bone Circuits



Ground-to-ground voice communications for the FAA are provided by the intercom/interphone (IP/IC) system, known as Service F. The interphone network is used primarily to support flight planning activities and flight movement (See Figure 7.3). The intercom network is utilized primarily for the transfer of ATC and flight related messages within and among FAA facilities. In addition, IP/IC is used for assistance and maintenance calls.

Aircraft-to-ground communications in NAS are basically through VHF voice channels. Aircraft-to-ground communications occur between an aircraft and the Flight Service Stations (FSS) and by commercial airlines on their own communication systems. The terrestrial VHF system extends over the whole country including 1,600 sites in the airline-operated segments (Figure 7.4).

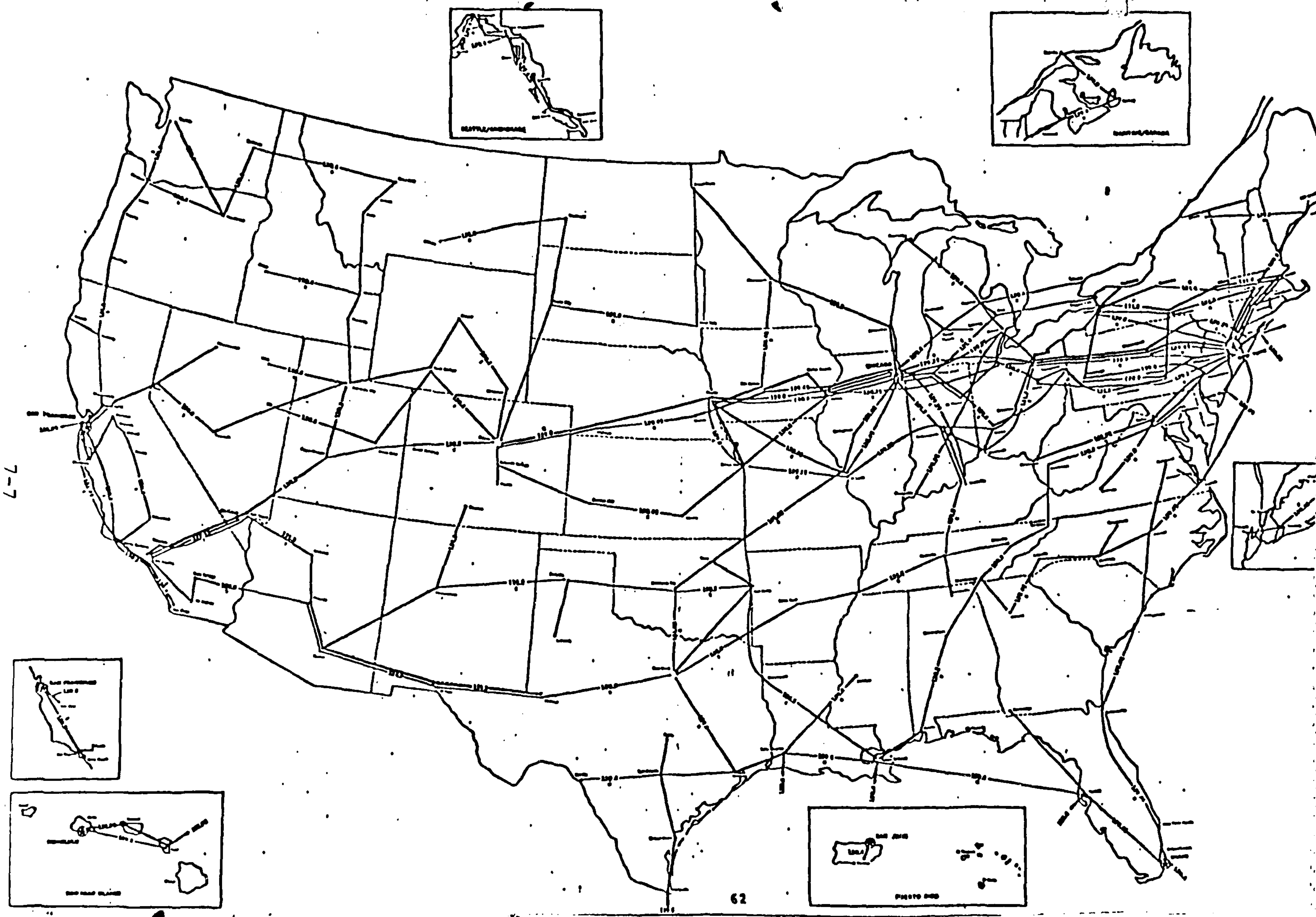
While the basic functions of NAS are assumed to not change, the system performance requirements of the NAS are likely to be more stringent. In particular, future systems will embody refined standards with respect to safety, reduced delays, collision avoidance, increased terminal capacity, better meteorological forecasts, etc. Voice communications between aircraft and the ground will continue but the frequency of use will diminish as automated data links become available and used. Further, the availability of efficient and economic digital communication systems will reduce the need for dedicated special design communication networks for civil aviation facilities.

Service F Interphone System Circuits

Outside Or FAA Facility To/From	ARTCC	ATCT	ATCT/ TRACON	FSS	CS/T
ARTCC	X	X	X	X	X
FSS	X	X	X	X	X
CS/T	X	X	X	X	X
ATCT (private or public)	X	X	X	X	X
ATCT/TRACON	X		X	X	X
RAPCON (IFR)	X	X	X	X	
RATCC	X	X	X	X	
ARAC	X				
BASOPS	X		X	X	X
WSO	X	X	X	X	X
ADC	X				
SAR(USCG)	X				
FAA-SCC(CARF)	X				
OVERSEAS LINK (ICAO)	X				
AUTOVON	X				
SOCS	X				
ARINC	X				
AIRLINES	X		X		X
PBX/UAS	X		X		X
FTS	X				
CENTREX	X				
PILOT BRIEF LOUNGE				X	
PRIVATE AVIATION		X			
ENTRANCE DOOR		X	X		
FAS		X			
KEY EQUIPMENT		X			
FAA-SCC/(CARF)			X		
ANG			X	X	X
FIREHOUSE			X		
GCA			X		
FX				X	X
COMMERCIAL					X

Figure 7.3

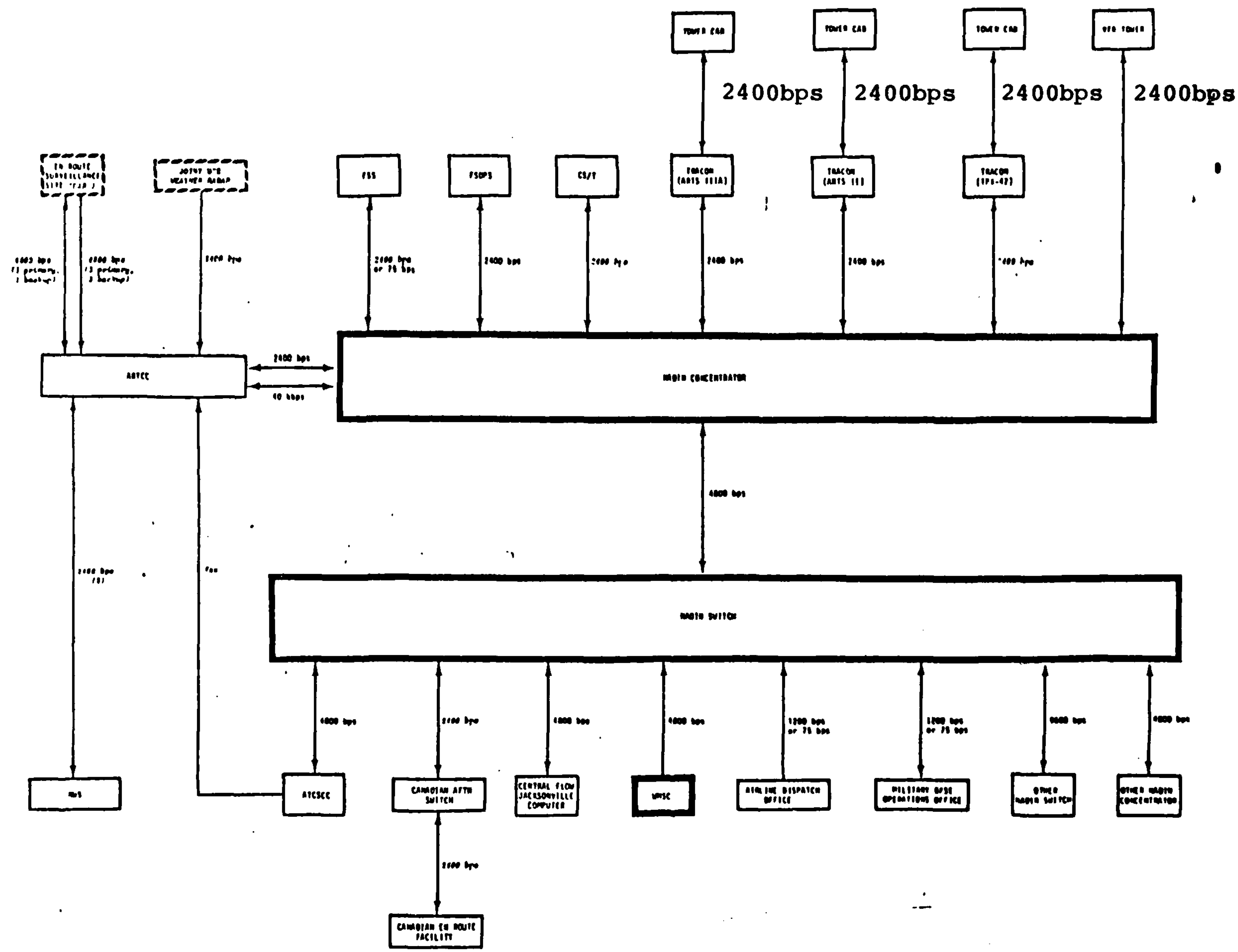
Airlines VHF Communications System Links



The aviation communication system for 2020 will be a function of available technology and the nature of navigation and surveillance deemed necessary by the regulatory authorities. Satellites will be an important element of any strategy to develop civil aviation communication. For example, the Department of Defense plans to change the military aviation navigation system to a satellite-based global system, the NAVSTAR Global Positioning System (GPS). If GPS or another satellite navigation/surveillance system is employed for civil aviation, the need for terrestrial ground based communication links among facilities may be eliminated. Thus, the ultimate configuration of the civil aviation communication system in 2020 will depend on the extent to which navigation and surveillance are terrestrially or satellite-based.

The next section briefly describes GPS, AEROSAT and other satellite based air traffic control (ATC) concepts and programs. A recent FAA study described the air traffic control system likely to evolve during the next decade (2). The projected 1980 NAS data communications system is shown in Figure 7.5. The proposed system assumes that the three phases of the NADIN program will be completed. NADIN will provide a more efficient data communications system than is available at present. In particular, NADIN will allow any user to communicate with any other user through the use of data concentrators and the two centrally located switches (Figure 7.2). In the 1990 time frame NADIN will be capable of

Project 1990 Data Communications System
for the National Airspace System

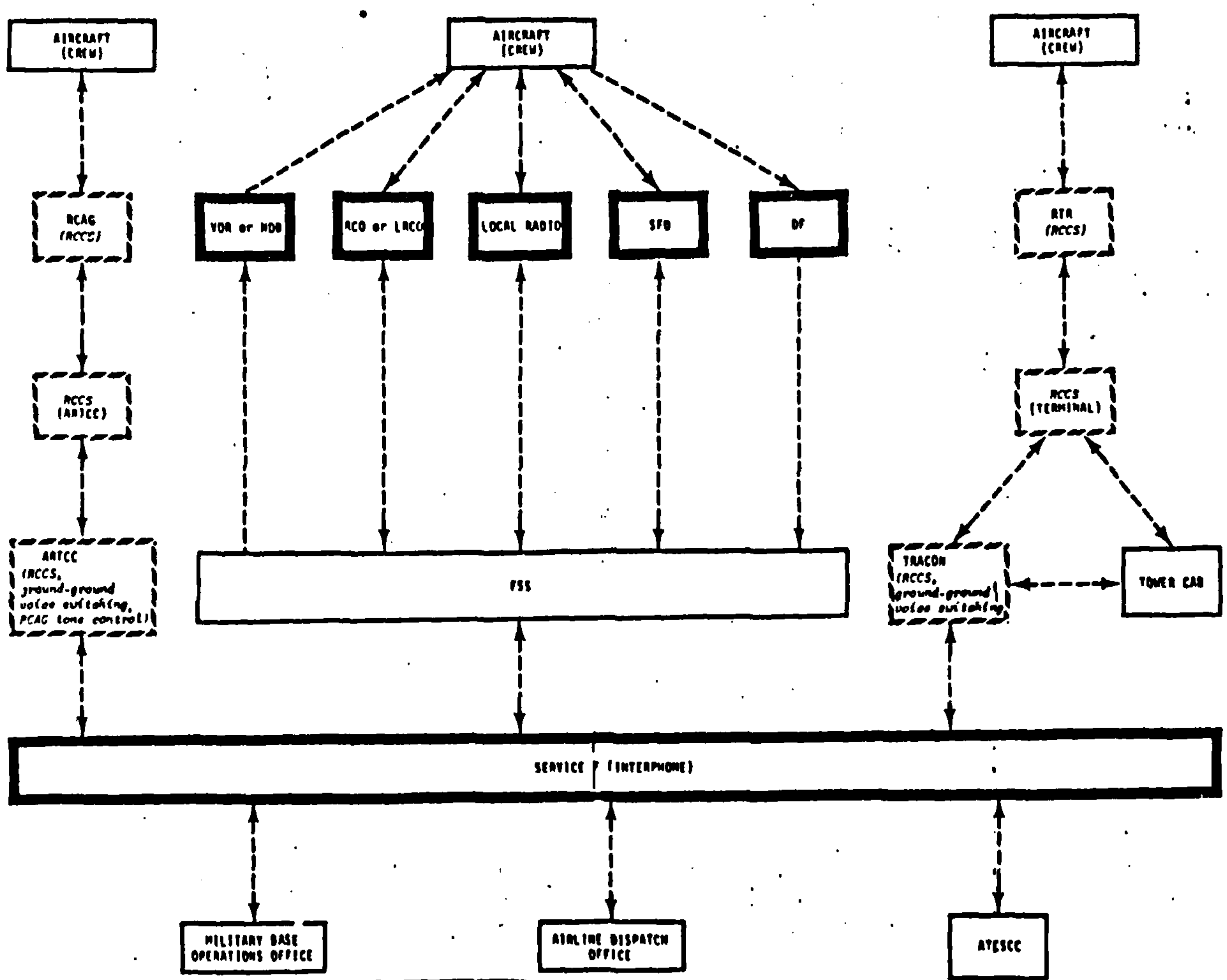


Source: FAA-EM-78-16

Figure 7.5

providing the ARTCC-ARTCC connectivity without the computer B network as backup. In addition, a 2300 bps data link will be established connecting centers and terminals. Aviation weather and flight service data dissemination will also be provided through NADIN. The projected 1990 NAS voice communication system is shown in Figure 7.6 (2). Both voice and data communications requirements are a function of the enroute surveillance and navigation systems and the terminal ATC requirements. Significantly different communication requirements, however, will result if the current terrestrially-based system evolves to a highly centralized space-based system.

Projected 1990 Voice Communications System
for the National Airspace System



Source: FAA-EM-78-16

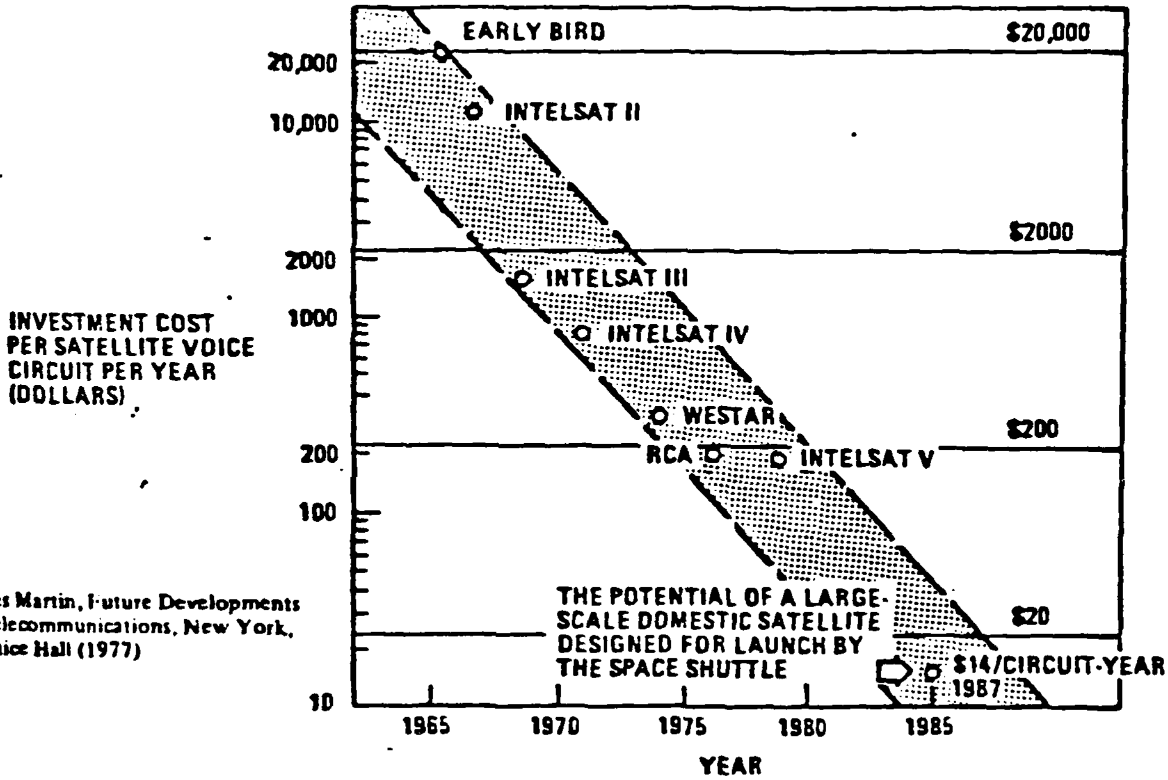
7.2 Satellite Systems

Satellite systems offer significant benefits to civil aviation. These benefits will be witnessed in navigation, surveillance and aircraft communications. In particular, satellites will allow:

- the provision of services simultaneously over a greater geographic area;
- the improvement of the quality of communications;
- the improvement of the accuracy in position reporting for aircraft;
- the provision of an independent altitude report for aircraft; and
- the provision of a uniform time basis for navigation and surveillance.

Thus far, the use of satellites for air traffic control has been constrained by both economic considerations and the uncertainty attendant to a new application of an existing technology. However, the continued use of satellite systems in a variety of communication and navigation functions will facilitate adoption for civil aviation. Further, continued use will allow the operating economics of satellites to become competitive with terrestrially-based systems. Satellites will be common and accepted technology by 2020; communication satellites will have been in service 55 years (see Figure 7.7) and the GPS system will have been in operation nearly 25 years.

Investment in Satellite Voice Circuits



James Martin, Future Developments in Telecommunications, New York, Prentice Hall (1977)

Figure 7.7

A geosynchronous communication satellite (NASA's AT-6) with three axis stabilization and a 30' antenna was first demonstrated in 1974. Larger geosynchronous satellites with multiple beam antennas of 100' to 200' diameter should be demonstrated by 1990. MARISAT, the satellite voice/data communication system for maritime users, launched in 1976, will have been in existence for 44 years by 2000. INMARSAT, the international group recently created to carry out the MARISAT function will be 40 years old.

Satellite based Search and Rescue (SAR), to be demonstrated on the next TIROS-N polar orbiting satellites, will have become operational in the 1980's. The SAR system may be expanded to integrate other aviation uses of low-orbit satellites, such as remote monitoring with data collection platforms of inaccessible naviads, marker beacons, etc. The anticipated relatively low operating and maintenance costs for satellites relative to ground based system will facilitate use. In addition, satellites will facilitate interconnection and monitoring of all facilities.

The global capability of satellite systems will allow significant improvement in standardization of user requirements in all parts of the world. Avionics for ocean and land areas would be congruent. At present satellites prove less expensive than ground based systems for high density point-to-point communications for distances in excess of a few hundred miles.

The costs of satellite communications will experience further reductions as low cost earth terminals proliferate in the 1980's. Such earth terminals will use KU-band of the Satellite Business Systems (SBS) and Advanced Westar System of Western Union domestically, and the Intelsat V series internationally (see Figure 7.8). During the 1990's, the technology will use the 20/30 GHz band. Any user group dispersed throughout the Continental United States or internationally, such as civil aviation, will find satellite services to be highly cost-effective for ground-to-ground communcations. These services will be offered commercially, and through data communication networks such as NADIN. In addition, the Weather Typewriter Services are likely to be provided via satellite, except for some high density short distance links.

The merits of satellite systems for ATC have been suggested in many studies. Concepts have been developed for both domestic and international civil aviation. The 1973 Advanced Air Traffic Management Systems (AATMS)* recommended that a ten-year research and development program be initiated to support a future decision on whether or not to implement a fourth-generation ATC system by the 1990's that would include satellite and ground-based elements. In addition, the FAA supported concept development of ASTRO-DABS, a satellite based position determination and data link system.

Trend in Satellite Earth Terminal Costs

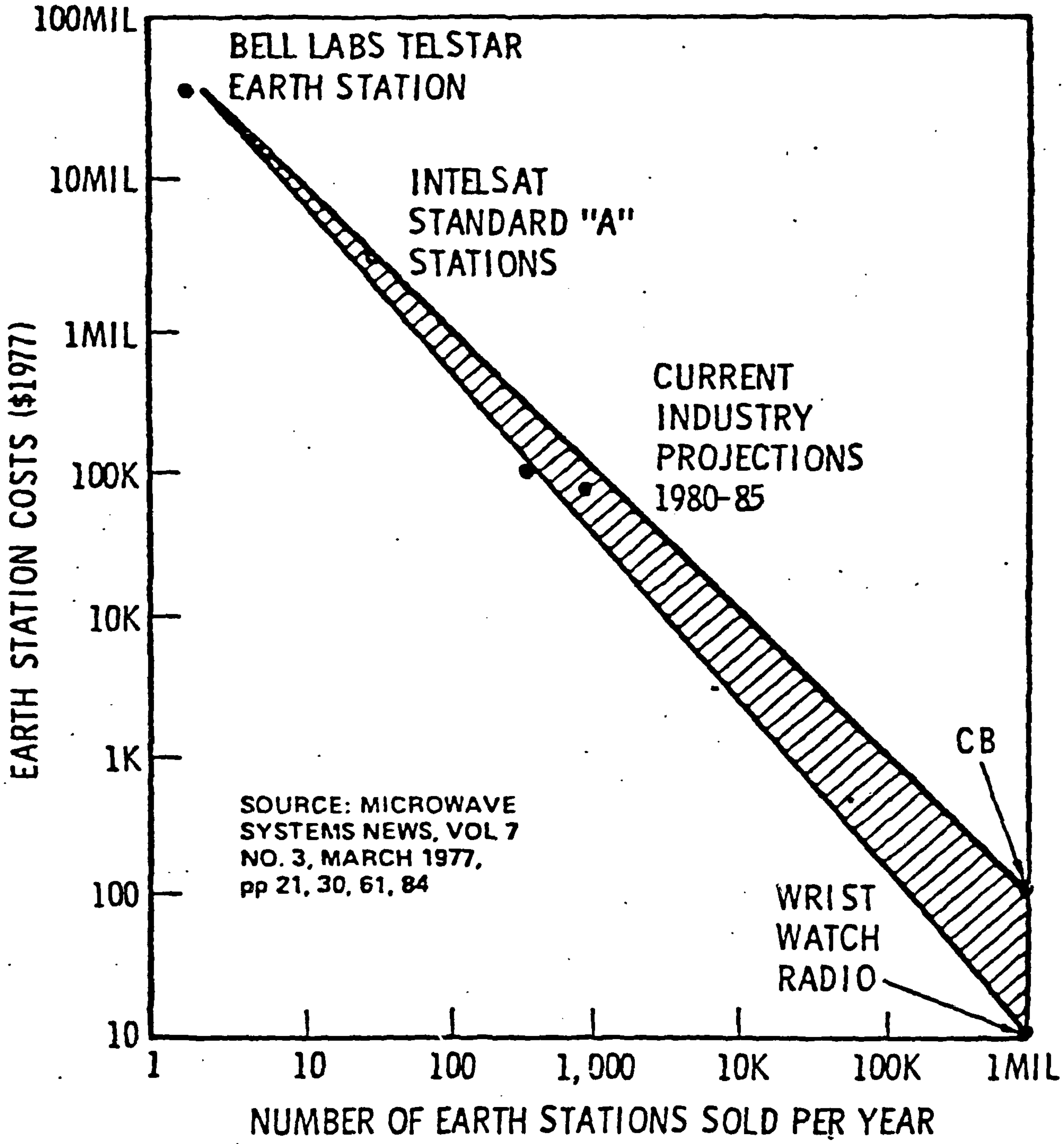


Figure 7.8

Similarly, the FAA participated actively in the planned AEROSAT program, a satellite-based oceanic system covering the Atlantic, Pacific and Indian Oceans. The MITRE Corporation, under the auspices of the FAA, has extended the previous studies of satellite based aviation systems, incorporating the GPS, with design details and tradeoffs. The study suggests that by the 1990's GPS and other satellite systems will be operational. By the year 2020, satellite systems for civil aviation will be in their second, if not third, generation of design. Some of the basic parameters of these various concepts and programs will be reviewed below.

a. <u>Overview of the Global Positioning System</u>

The Global Positioning System, also called NAVSTAR GPS, is designed to provide signals for navigation. The GPS does not have communications capability. The system will provide accurate position determination, velocity, and system time to both military and civil users all over the world. The provision of such data requires the processing of psuedo-random frequencies transmitted at L-band from four satellites. By design, the accuracy of the information available to the civil users will be less than that provided the military user. Implementation efforts for the system are proceeding as shown in the plan presented in Figure 7.9. Experimental results indicate that the system is achieving the level of accuracy defined in the project objectives.

Overview of GPS Program

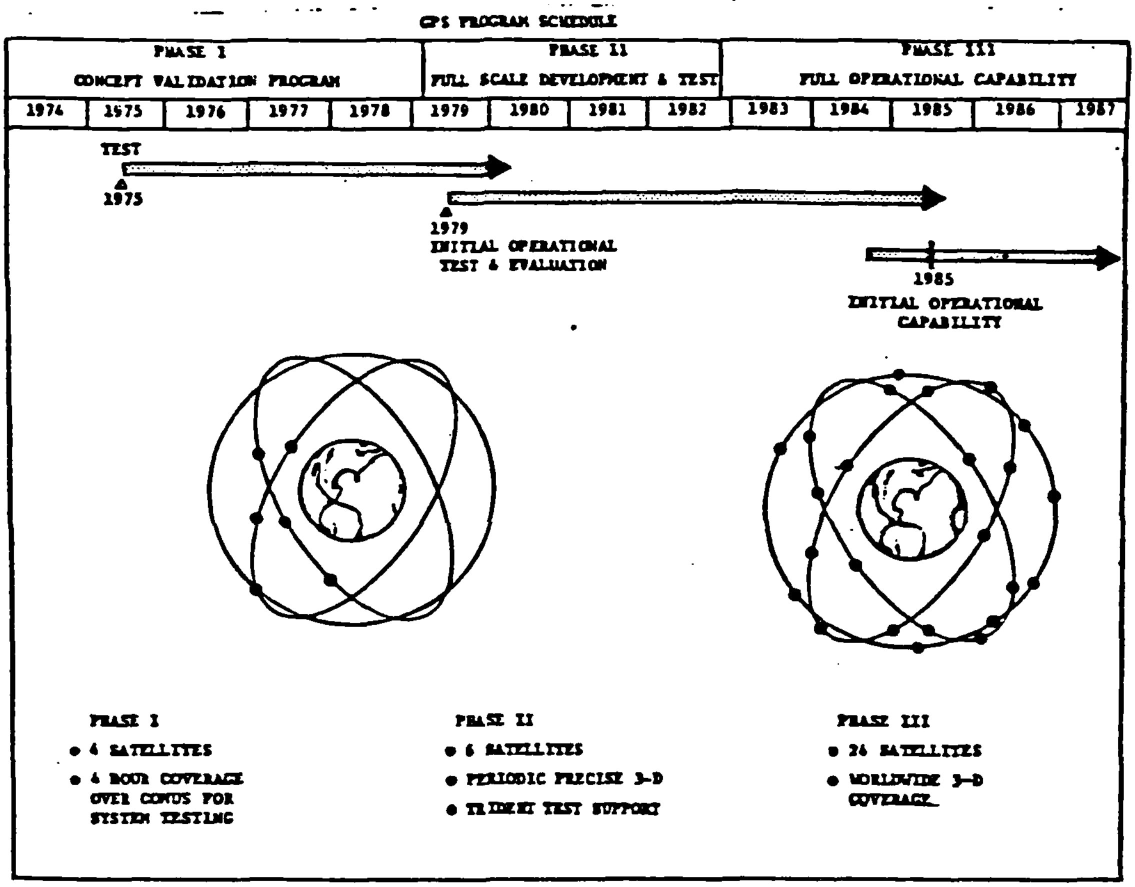
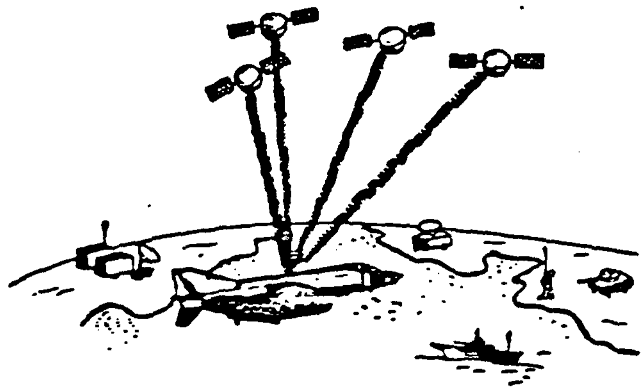


Figure 7.9

In the operational phase, the space segment will consist of 24 satellites. The satellites will be placed in 10,000 nautical mile (n. mi.) circular orbits. A satellite will complete an orbit every 12 hours. Eight evenly-spaced satellites are to be arranged in each of three orbital plans, includes 63^{o} and spaced 120^{o} apart. Each satellite will transit a precision (P) signal and a coarse/acquisition (C/A) signal on two L-Band carriers, L_1 and L_2. The P signal will be available only to authorized users. The P signal is designed for very accurate positioning performance (10m,2 σ) and will provide resistance to jamming, spoofing, and multipath. The C/A signal, to be available to both military and civil users, is designed for less accurate positioning performance (100m, 2 σ) and as an aid to acquiring the P signal. Both signals are to be continuously transmitted in phase quadrature on L_1, while only one is to be transmitted on L_2 at a given time. In the present design, L_2 is reserved for the P signal. The satellite navigation subsystem is displayed in Figure 7.10.

The ground segment of GPS will perform satellite tracking and control functions as well as determine the navigation data (e.g., ephemerides, clock offsets, and signal propagation error data) to be superimposed on the coded satellite navigation signals. The ground segment will require several monitoring stations to provide adequate satellite tracking, a master control station, and an upload station. Good quality measurement data from each monitoring

GPS Satellite Navigation Subsystem Overview



CHARACTERISTICS

- SIGNALS: L_1 - 1575 MHz (P & C/A CODE)
  L_2 - 1227 MHz (P or C/A CODE)
- EXPECTED ACCURACY (90%)

	P SIGNAL	C/A SIGNAL
POSITION	30 FT	< 300 FT
VELOCITY	0.1 FT/SEC	< 1 FT/SEC
TIME	10 NS	< 100 NS

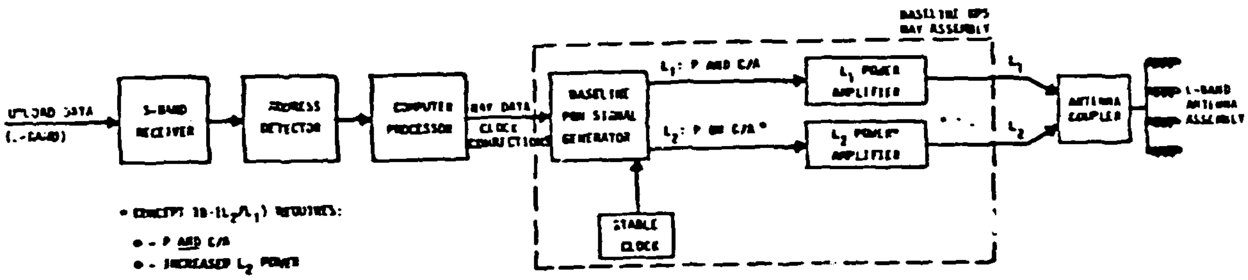


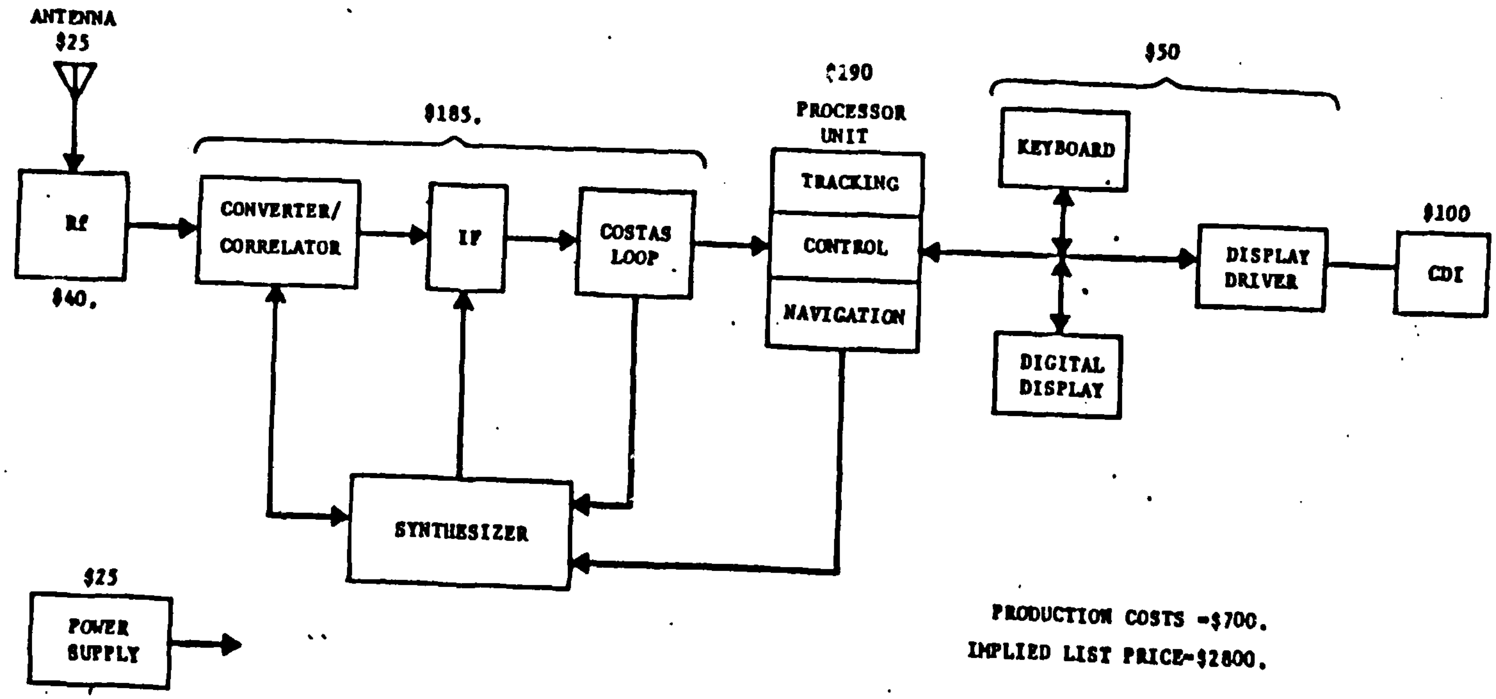
Figure 7.10

station will be processed at the master control station. The processed data will be navigation information for uploading at S band once per day to each satellite via the upload station.

The user segment of GPS consists of receiver/processor equipment for air, marine and land mobile users. The nature of the receiver/processor equipment will depend upon civil user and military requirements for navigation accuracy and jamming immunity. User equipment for most civil applications would probably be functionally similar to the military z set which sequentially acquires and tracks the C/A signal from the four satellites in the navigation process. A substantial effort is in progress to develop a low cost GPS receiver for civil applications since present cost estimates of the military z set are not economic for civil aviation. A recent study indicates that significant cost reductions are possible for a civil GPS receiver. A block diagram of such a receiver, with the projected 1985 costs, in 1977 dollars indicated, is shown in Figure 7.11. The civil GPS receiver costs are based on current trends in digital and integrated circuit developments and estimated production quantities of 5,000 units. The cost estimates are based on current avionics costs and may decrease proportionately concommitent with new technological developments. The cost for VOR, DME and LORAN units are shown in Figure 7.12.

Low Cost GPS Receiver Block Diagram

(1985 Avionics Cost Estimates)



PRODUCTION COSTS =$700.
IMPLIED LIST PRICE=$2800.

Figure [illegible] 11

1985 Cost of Avionics Used in Economic Analysis

GENERAL AVIATION	LIST PRICE
VOR	$ 900
DME	1,800
RNAV Computer for VOR/DME	1,000
LORAN Receiver/Navigator	2,050
GPS Receiver/Navigator	2,800

1977 Dollars

Figure 7.12

b. AEROSAT Program and Oceanic ATC

The AEROSAT program developed a detailed design concept and system implementation for oceanic ATC. An international working group with active FAA involvement concluded that a satellite-based system would offer the capability of providing large capacity, good quality, reliable communications and surveillance for large geographic areas. The system would provide surveillance for the oceanic areas now served only by HF communications including the airspace over the Atlantic, Pacific, and Indian Oceans. The primary service areas would include the North Atlantic, Caribbean, Central East Pacific, and Western Pacific. Development of the satellite system over the three major oceanic areas would provide capabilities for all the oceanic airspace as well as all the airspace over land masses, as shown in Figure 7.13.

The system concept includes the full range of services for air traffic control and airline company communications. This system concept also incorporates improvements to terrestrial ATC and airline communications systems.

The oceanic ATC and communication system concept is shown in Figure 7.14. The ATC/CSC consist of one or more automated oceanic ATC centers that is visible to each pair of satellites, and a communications subsystem (3). The geostationary satellites relay

Global Coverage Provided by a
Five Satellite Oceanic ATC System

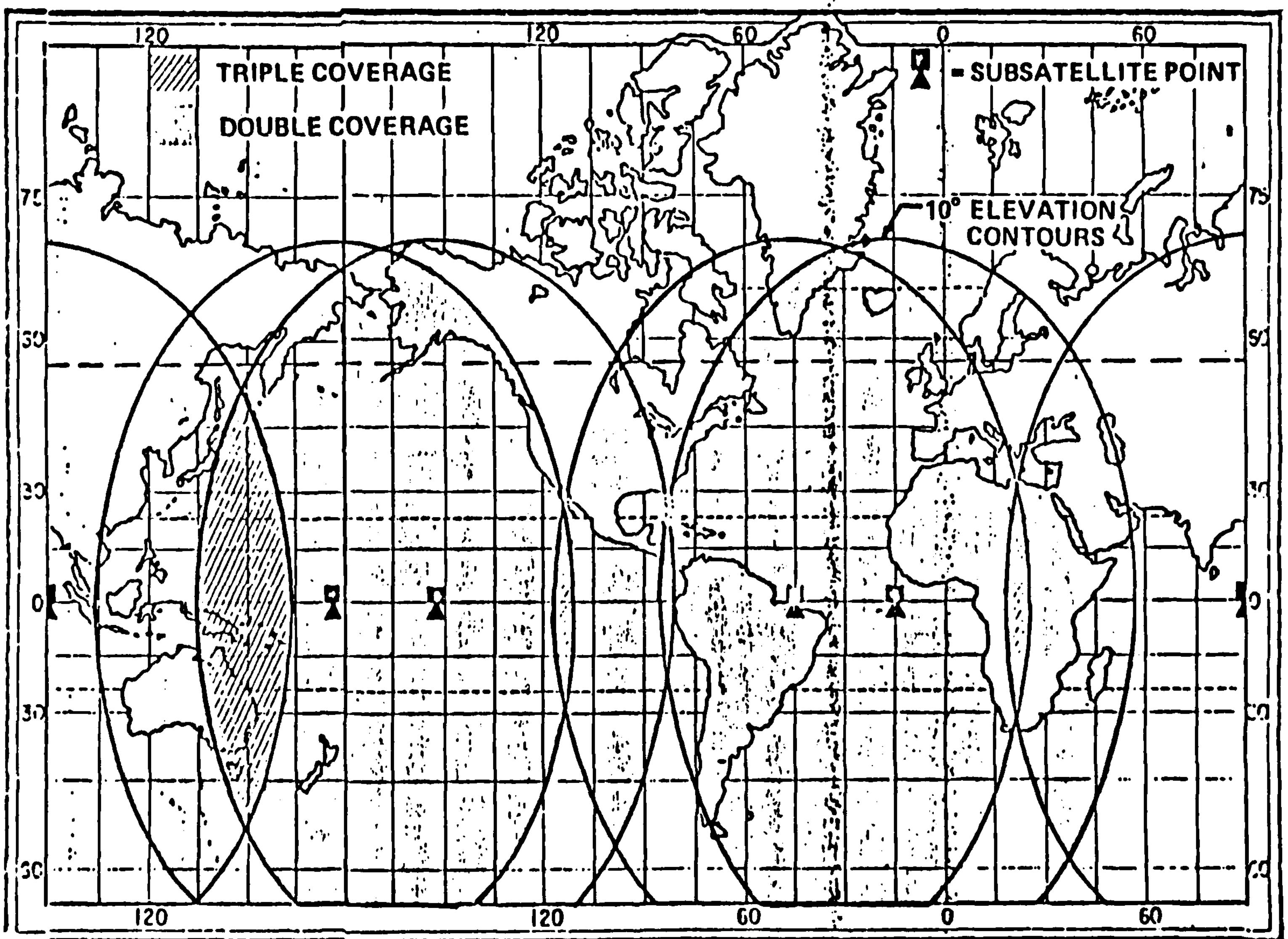


Figure 7.13

Oceanic Satellite-Based ATC Concept

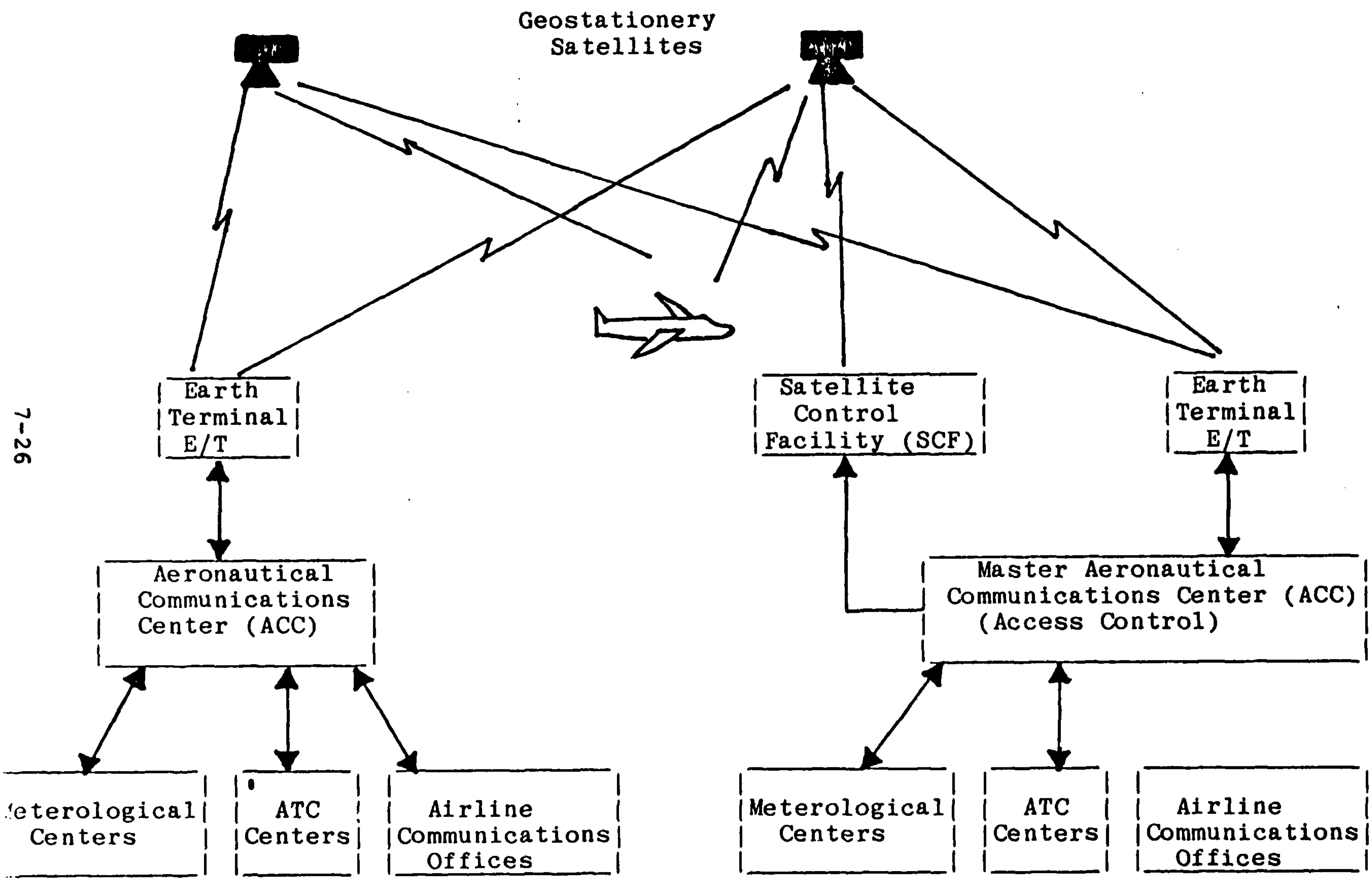


Figure 7.14

ground-to-aircraft (G-A), aircraft-to-ground (A-G) and ground-to-ground (G-G) communications. A number of earth terminals (E/T) would be employed to receive and transmit communication and ranging signals to and from the satellites. An aeronautical communications center (ACC) is associated with each earth terminal. The ACCs would be connected with the various ground users, such as the Oceanic and Enroute ATC Centers, meteorological centers, and airline company communication offices. Such connections would be accomplished using appropriate common carrier networks.

One of the Aeronautical Communications Centers serves as a Master facility as shown in Figure 7.14. This facility controls satellite access for all users. The Satellite Control Facility (SCF) is located also at the Master ACC. Data polls operated by the Master ACC provide the basic means of communicating A-G and G-A. G-A data communications from all ground users are relayed via satellite A-G channels, first to the Master ACC, which transmits them to the aircraft on the uplink of the data poll with an A-G data transmission, which can include the aircraft's position (dependent surveillance), and an A-G ranging response if equipped for independent surveillance. All ACCs monitor A-G transmissions and distribute communications as required to attached ground users.

Full duplex direct voice communication service is provided on a request basis. Voice channel requests from ground and airborne users in the form of fixed format data messages are relayed to the Master where a single queue for voice service is maintained. As voice channels become available, they are assigned to the airborne user via the data poll uplink and ACC serving the specified ground user via the G-G satellite link. Once channel assignments are made, the voice communication commences directly from the serving ACC to the aircraft.

Two types of surveillance concepts can be used. Dependent surveillance would be to simply relay the aircraft's own navigation system information to the ground stations, as mentioned above. Independent surveillance would require a ranging transponder on each aircraft. All aircraft would transmit a range code in response to a ranging signal, which is relayed via two satellites to the originating ground station, where aircraft position is calculated from two range measurements and altitude relayed from the aircraft altimeter.

The economic assessment of oceanic ATC indicated that the satellite system results in significant long term savings. Although the program was not initiated, due to various institutional difficulties and large initial investments required, the basic potential and interest in oceanic satellite ATC communications still exists, and it is likely that in the next several years

some elements of such a system will be in place. If indeed the system benefits are as indicated by the various studies, it can be expected that the Oceanic ATC plan of Figure 7.15, may be realized with a shift in time of 5 years. Significant progression over the last few years has been made in definition of AEROSAT by the COMSAT Corp. which was ready to initiate the program as operator for the U.S., Canada, and ESA.

c. Satellite Based DABS Concepts and Extensions

ASTRO-DABS is a Satellite-Relayed Surveillance, Navigation, and Data Link Concept studied by FAA for air traffic control in CONUS since 1972. The concept was examined in terms of its technical feasibility, performance, and cost-effectiveness. It was found that satellite ATC was feasible but the high costs of avionics, the threat of uplink jamming, the lack of an air-to-air CAS (collision avoidance system) mode, and the risks associated with advanced technology, significantly reduced the attractions of universal coverage and independent altitude measurement associated with satellite ATC[4]

ASTRO-DABS was later modified and extended to reduce the costs of avionics, to develop an air-to-air CAS mode, and to reduce the threat of jamming. The design goals of "ASTRO-DABS 1974" are summarized in Figure 7.16. The design concept has 7 satellites in synchronous equitorial orbits and 6 satellites in non-equatorial synchronous orbits, as shown in Figure 7.17.

Oceanic ATC Program Plan of 1975

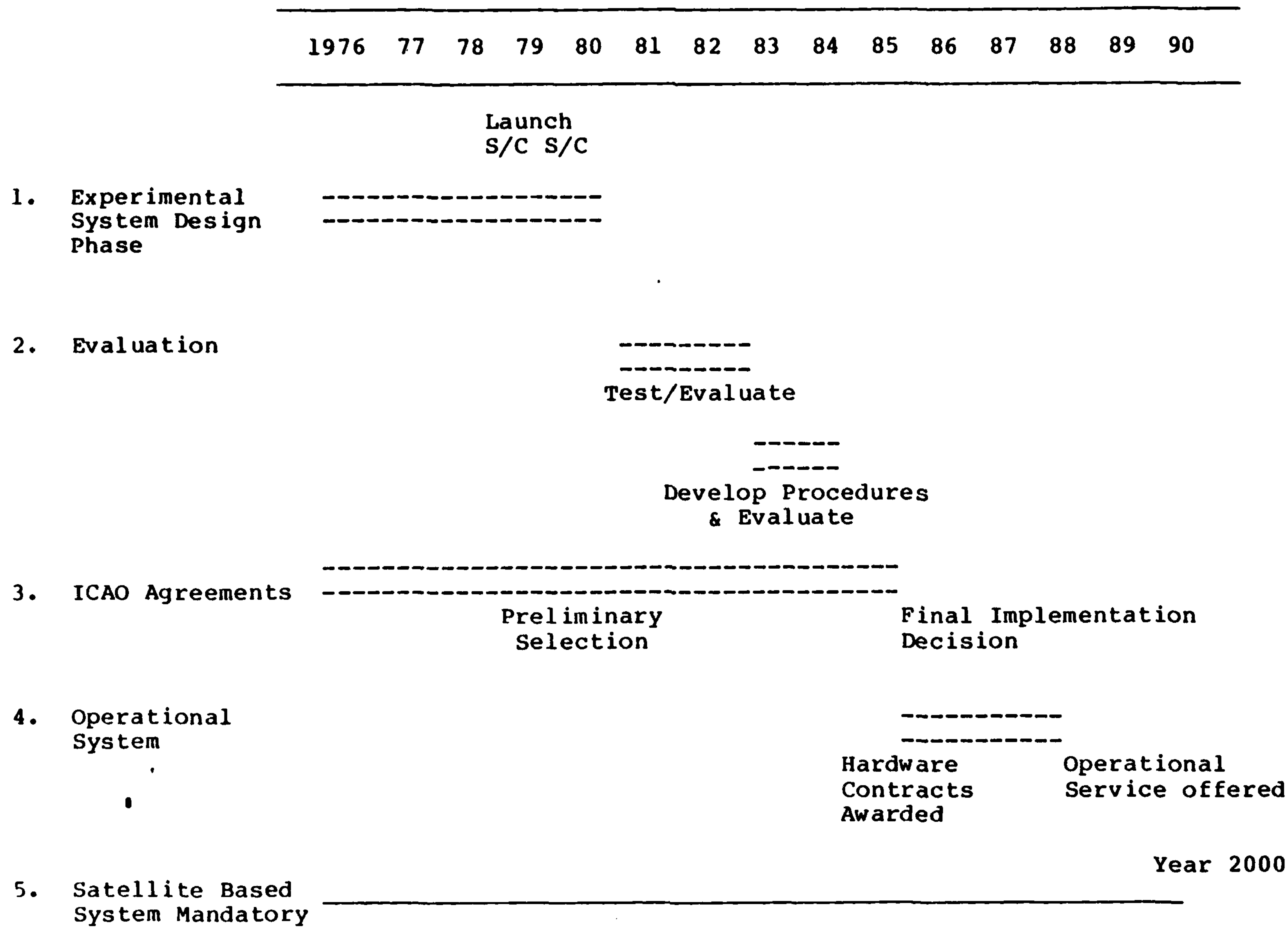


Figure 7-15

ASTRO-DABS 1974 DESIGN GOALS

SURVEILLANCE

Address Population	4,000,000
Instantaneous Airborne Capacity	50,000
Instantaneous Airborne Capacity Growth Capability	100,000
Average Interrogation Period	4-4 1/2 sec.
Altitude Estimate Uncertainty (1σ)	
Absolute	100 ft.
Relative-10 miles-distance between two aircraft	100 ft.
Lat/Long Estimate Uncertainty (1σ)	
Absolute	50 ft.
Relative-10 miles-distance between two aircraft	50 ft.

DATA LINK

Number of bits per Interrogation	
75% Surveillance Only	28 bits
20% Surveillance and Position Warning Indication (PWI) - 2 Target Capability	56 bits
4% Surveillance and Intermittent Position Control Commands (IPC) and PWI - 5 Target Capability	84 bits
1% ATC Data Link	84 bits
Probability of Bit Error	10^{-6}

NAVIGATION

Average Update Rate Exceeds	1/2 sec.
Altitude Estimate Uncertainty (1σ) - Absolute	150 ft.
Lat/Long Estimate Uncertainty (1σ) - Absolute	75 ft.
Steady-State Position Update Computation Time	milliseconds

AIR-TO-AIR CAS TARGET POSITION DETERMINATION

Average Update Rate	4-4 1/2 sec.
Altitude Target Estimate Uncertainty (1σ) - Relative 5 miles	100 ft.
Lat/Long Target Estimate Uncertainty (1σ) - Relative 5 miles	50 ft
Position Update Computation Time	milliseconds

Figure 7.16

ASTRO-DABS Constellation

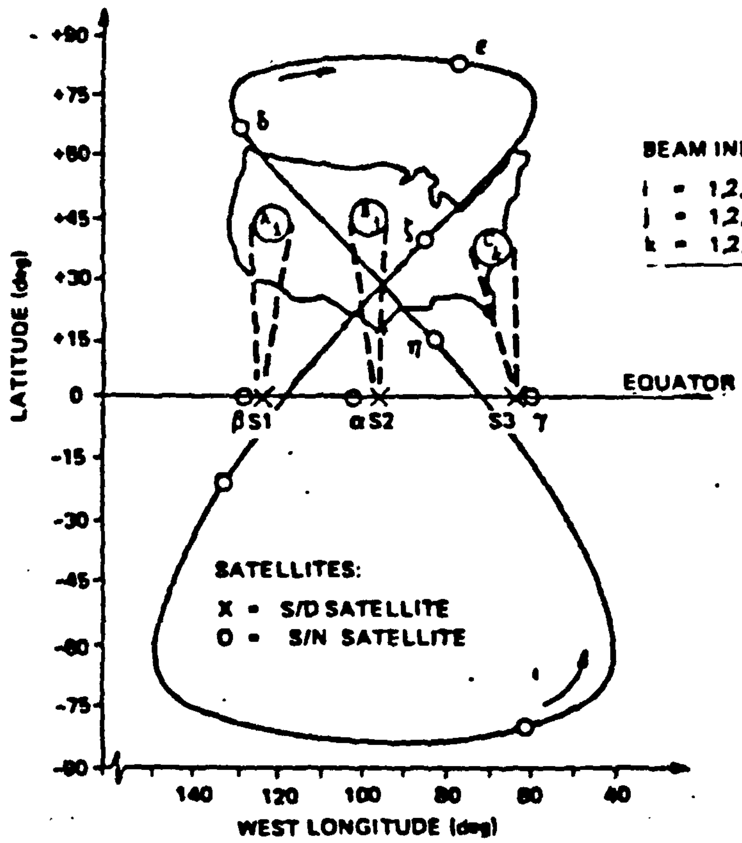


Figure 7.17

ASTRO-DABS was designed to handle a peak instantaneous airborne capacity of 50,000, with surveillance updates to each aircraft of 4 to 4 1/2 seconds. The surveillance position uncertainty was very small, with absolute aircraft uncertainty estimated to be not more than 100 ft. in altitude (1 σ and relative to one point in CONUS). The uncertainty in the relative altitude between two aircraft located 10 miles or less from each other was also not more than 100 feet (1 σ). Horizontal aircraft position, both absolute and relative has half the uncertainty that altitude had, due to geometric considerations, thus a 50 ft (1 σ) value was predicted.

The data link was designed to handle all IPC (Intermitent Positive Control) link messages (comparable to the DABS data link with a bit error probability of no more than 10^{-6}.

ASTRO-NAV was designed so that the update rate exceeded 1/2 second with the uncertainty in altitude being on the order of 150 ft (1 σ) and 75 ft (1 σ) in latitude and longitude, respectively. The slightly degraded performance in ASTRO-NAV, over that which occurs in ASTRO-DABS surveillance was due to differences in clock stabilities assumed for the Satellite Control Center (SCC) and the aircraft.

ASTRO-DABS has a unique three-dimensional air-to-air CAS which can provide each aircraft with very accurate three-dimensional positional information of all targets in its proximity. The predicted relative accuracy of the air-to-air CAS mode is comparable to that of the surveillance mode of ASTRO-DABS.

To reduce the avionics costs, an integral surveillance backup mode has been designed which eliminates the need for an encoding altimeter to interface with a ground-based DABS site. The surveillance backup transmissions are sent directly to the ground so that they are not vulnerable to uplink jamming. The accuracy of the surveillance backup aircraft position estimates is predicted to be almost equivalent to the accuracy estimates generated by the primary surveillance mode of ASTRO-DABS.

In the original ASTRO-DABS concept, the system could double its surveillance capacity by filling in the gaps between transmissions with a second set of range ordered transmissions. Separability was achieved by having separate aircraft-to-satellite links. Thus, with doubling of satellite average power and an increase in bandwidth at L-Band of 10 2/3 MHz, ASTRO-DABS could double its capacity with no penalty to user equipment and no need to provide more than two aircraft-to-satellite frequency channels.

In ASTRO-DABS with air-to-air CAS, the doubling capacity (100,000) is realized just as it is in the original concept, except that the number of CAS channels will also have to be increased from two to four. Again the IPC and surveillance-only users are not penalized when the capacity of the system is doubled; however, care has to be taken not to penalize the CAS-equipped user. This can be achieved in the original design by giving the CAS equipped-user a four channel capability initially, even though he may never use more than two channels, or by designing the CAS receiver to allow for a relatively expensive modem add-on, should a four frequency capability become a requirement.

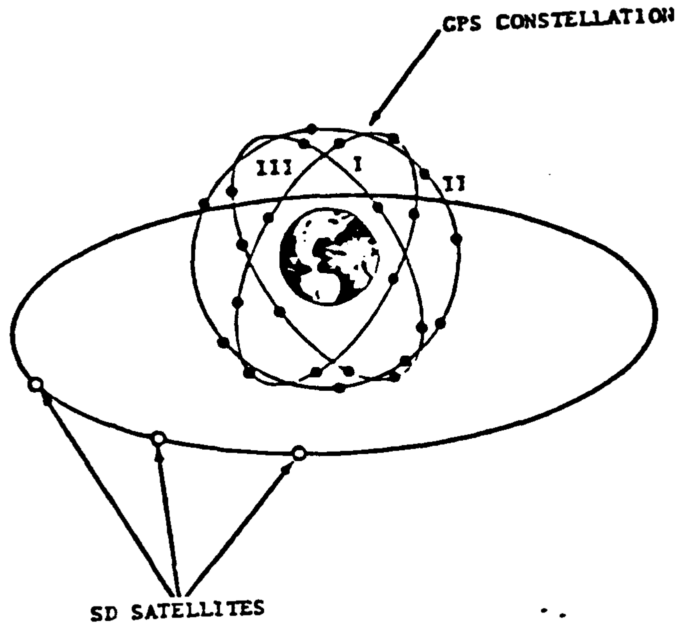
The GPS program has resulted in further concept development of satellite-aided surveillance and data link systems with GPS serving as the navigation element. This work is reported by Elrod et al. [5] and it investigates several satellite and terrestrial system concepts for implementation in the late eighties with traffic levels for the 1990s.

This study considers both the use of GPS and modifications to the GPS signal for low cost civil use. Also considered are additions of repeaters on each GPS satellite to provide an independent space-based surveillance system -- although the space platform would be the same for both. Communication data link added to the above GPS based surveillance and navigation system is considered

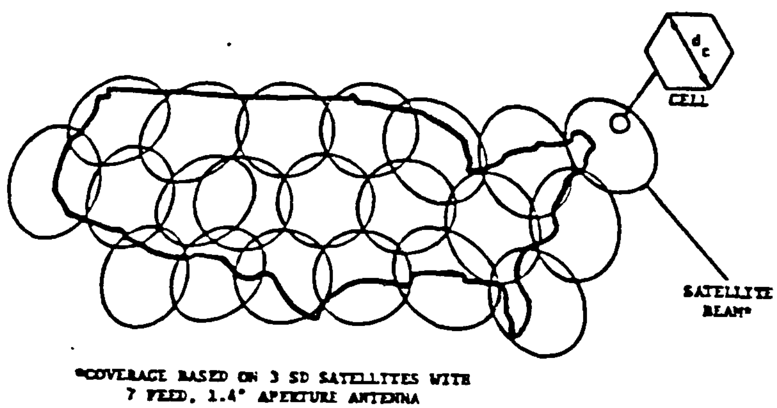
for both space and terrestrial approaches. To add surveillance and data communications geosynchronous equatorial satellite relays as employed in ASTRO DABS, the resulting satellite constellation is shown in Figure 7.18. This system design is for 1995 CONUS requirements, detailed in Figure 7.19.

The synchronous surveillance and data link satellite considered in this concept are of the NASA ATS-6 type. Three of these satellites are employed to meet the 50,000 Instantaneous Airborne Count (IAC) forecast for 1995. Each satellite has a 5MHz bandwidth and 7 beams, resulting in 21 beams over CONUS. Note that the AEROSAT Oceanic ATC system design was also based on the ATS-6 thirty feet diameter antenna.

Overview of Surveillance and
Data Communications Satellites
and GPS



a. Satellite Constellation



b. Surveillance-Data Link Satellite Beams

Figure 7.18

1995 Performance Requirements for CONUS ATC System

ITEM	SURVEILLANCE	NAVIGATION*	DATA COMM. (A/G)
CAPACITY	50,000 IAC	UNLIMITED	50,000 IAC
UPDATE INTERVAL	4-6 SEC	0.5 SEC	4-6 SEC
TIME TO FIRST FIX	-	<60 SEC	-
ACCURACY (95%)			
- TERMINAL (HOR/VERT)	100 FT/200 FT	1000 FT/200 FT	10^{-6} BER
- EN ROUTE (HOR/VERT)	400 FT/200 FT	3000 FT/200 FT	10^{-6} BER

*INCLUDES NON-PRECISION APPROACH CAPABILITY

BER - Bit Error Rate
IAC - Instantaneous Airborne Count

Figure 7.19

7.3. Requirements and ATC System Configurations for the Year 2020

a. Aviation Activity Forecasts

Requirements in the year 2020 for aviation communications will be a function of the socio-economic developments in the next forty years. In Phase II, three contrasting scenarios were developed as follows:

- Balanced Growth - Scenario X
- Rapid Growth - Scenario Y
- Stagflation - Scenario Z

Figure 7.20 lists the forecasted aviation activity for the 3 scenarios. This table includes the important variables for sizing the NAS requirements. The total fleet size is used for estimating user costs, in which general aviation dominates. The peak IAC at major hubs defines the capacity of systems that have coverage areas of the size of a major hub or smaller area. The peak CONUS IAC defines the capacity that systems with CONUS size coverage need to satisfy (e.g., satellite systems).

The 1995 aviation activity forecast employed in earlier studies is included in Figure 7.20. The data presented here is from the FAA's Advanced Air Traffic Management System B Study, carried out by MITRE Corporation.[8] That model has since been further refined; however, for the purpose of this exercise, the modifications are minor.

AVIATION TRAFFIC FORECASTS

Scenario	Number of Aircraft			Peak IAC			
				Major Hubs			CONUS
	GA	AC	MIL	GA	AC	MIL	
Rapid Growth	680	N/A	N/A	5,700	370	67	100,000*
Balanced Growth	510	N/A	N/A	1,900	154	19	60,000*
Stagflation	260	N/A	N/A	1,450	178	22	30,000*
MITRE AATMS Projection for 1995	335	5	20	1,673	128	39	37,022

N/A = Not available in Phase II Study
* Assumed in proportion to Peak Hub IAC and National Operations

Figure 7.20

In the Phase II report, the numbers of air carrier and military aircraft were not tabulated, and hence are not shown in Figure 7.20, nor was the peak CONUS-IAC derived. Using a very informal procedure for the purpose of this study, the peak CONUS-IAC has been estimated as indicated. Slight variations from these assumptions will not affect the basic system concepts to be developed.

It should be noted that the year 2020 forecast activity for the stagflation scenario is less than the 1995 AATMS model, which is reasonable since the basis of the earlier 1995 model was an equivalent "rapid growth" scenario from 1973-1995[8]. On the other hand, the traffic forecast for the 2020 rapid growth shows a tripling of peak hub IAC compared even to the balanced growth forecast. Such a large traffic increase would require significant system capacity and performance increases. To obtain this improved capability in a cost effective manner will require significant automation and structural changes to the current National Airspace System.

Future CONUS requirements for air-ground communications for the 1995 time frame are addressed in a recent RTCA report[9]. For the purpose of this study these same requirements, slightly extended, would be appropriate. Digital air-ground communications requirements are subdivided into short ATC messages and long ATC messages. Short messages are estimated to have a maximum of 30 characters per message. Figure 7.21 lists the short message types and Figure 7.22, the long messages types from Reference [9].

Typical Short ATC Message Types
(Short Messages less than 30 characters/message)

SHORT ATC MESSAGE TYPES	DATA FIELD
Altitude Assignments	
Climb Now to and Maintain (Alt) Descend Now and Maintain (Alt) Maintain Your Present Altitude of (Alt) Restrict Altitude Now to Between (Alt_1-Alt_2)	Zero to 99,900 ft. in 100 ft. increments
Heading Assignments	
Turn Right Now to and Maintain (Deg) Turn Left Now to and Maintain (Deg) Maintain Your Present Heading of (Deg) Restrict Heading Now to Between (Deg_1-Deg_2)	Zero to 360 degrees in 1 degree increments
Speed Assignment	
Increase Speed Now to and Maintain (Kts) Decrease Speed Now to and Maintain (Kts) Maintain Your Present Speed of (Kts) Restrict Speed Now to Between (Kts_1-Kts_2)	Zero to 999 knots in 5 or less knot increments
Voice Frequency Assignment	
Switch Now to ATC Voice Channel (Freq) Contact ATC Now on (Freq)	VHF or UHF in 25 kHz increments
Altimeter Setting Assignment	
Set Altimeter Now to (in. Hg or millibars)	4 digits
Traffic Advisory	
Relative Bearing (12 bearing sectors) Relative Altitude (above, at, or below recipient)	

Figure 7.21

Average Mess[illegible] Rates f[illegible] Short A[illegible]
Messages for High Density Hub in 1995
(Short messages are of the order of 30 characters/message)

Aircraft Category	UPLINK Per Aircraft Message Rate (Messages/Aircraft/Min)	UPLINK Number of Aircraft	UPLINK Average Message Rate (Messages/Sec)	DOWNLINK Average Message Rate (Messages/Sec)
IFR Arrivals	1.09	67	1.2	
IFR Departures	0.70	67	0.8	
IFR Holds	0	16	0	Pilot Acknowledgement
IFR Overflights	0.26	285	1.2	
IFR Withins	0.32	45	.3	
VFRs in TCA and on Reserved Routes	0.30	265	1.3	
ATC Subtotal			4.8	4.8
IPC in Mixed Airspace				
Advisories	6.0	1,010	101.0	
Commands	0.80	1,010	13.5	13.5
VFRs in TCA and on Reserved Routes	0.10	265	.4	
IPC Subtotal			114.9	13.5
Pilot Initiated Requests	—	—	—	0.4
Total Messages/Second (ATC and PIC Combined) for 1,700 Aircraft			119.7	18.7

Figure 7.22

The resulting short message rate is given in Figure 7.23. Further definition of these digital air-ground requirements is necessary; however, for this work the above description is given to indicate the significnt transfer to digital communictions of current analog communications. The digital ATC communications requirements above are based on a MITRE study [10], and this study was employed in the system concept definiton of reference [4].

Airline air-to-ground communications requirements are detailed in Figure 7.24 from reference [9]. This table shows a large potential for digital air-to-ground communications and a fairly low voice communication requirement. Further analysis is necessary to determine the exact number of air-to-ground voice communications required by the year 2020. However, future voice communication requirements are expected to decline drastically to perhaps one half of today's level, with increased availability of digital data link for ATC communications and with major advances in projected cockpit and ground station terminals and displays. These requirements will be further discussed in terms of the various future system concepts.

Character Rates for Long ATC Messages

Figure 7.23

ICAO MESSAGE TYPE ICAO MESSAGE TY	UPLINK OR DOWNLINK	ICAO MSG. RATE ESTIMATE	ICAO CHARACTERS/ MESSAGE	EXAMPLE: 1 HOUR FLIGHT		
				MSGS/ FLIGHT	CHARS/ MESSAGE	CHARS/ FLIGHT
ATC Messages (Estimates apply to all IFR aircraft)						
Flight Plan Revision	D	V	V	.2	100	20
Clearance Revisions	U	V	V	.2	100	20
Non-Control Advisories (e.g., ATIS)	U	V	300	1	300	300
Ground to Air Weather (En route)	U	V	500 Max.	1	250	250
Ground to Air Weather (Terminal)	U	V	250 Max.	1	125	125
Terminal Update (ATIS)	U	V	16 to 100	1	50	50
					TOTAL =	765
			Average Characters/Aircraft/Minute:			12.3

V = Variable

Airline Message Requirements for Average Day - 1980

Message Categories	Digital Messages			Voice Messages		
	Data Chars. per Msg.	Average Contacts per Flight	% of Flights Utilizing Service	Voice Secs. Per Msg.	Average Contacts Per Flight	% of Flights Utilizing Service
A. FLIGHT OPERATIONS						
1. Departure/Arrival Reports	69	3	91.5	20	3	8.5
2. Enroute Position Reports	69	1	35	24.5	1	8.5
3. Enroute Flight Plan Revisions	495	1	5	49.5	1	1.2
4. Flight & Ground Delay Reports	25	3	6	25	3	2
5. Weather & Aircraft Environmental Reports	400	1	35	49.5	1	6
6. Gate Assignments	10	1	80	10	1	10
7. Crew Physiological & Performance Monitor	15	4	100	—	-	-
8. Manifest Check (weight & balance)	60	1	75	25	1	10
9. Miscellaneous	300	1	20	30	1	5
B. IN-FLIGHT MAINTENANCE SUPPORT						
1. Airframe/Engine Parameters	15	4	100	—	-	-
2. Engineering Assistance	-	-	-	115	1	6.2
3. Miscellaneous	300	1	5	30	1	1.2
C. LOGISTICS/CUSTOMER SERVICE						
1. Seat Occupancy, Flight Connections/etc.	820	2.5	100	—	-	-
2. Immigration, public health clearance	5000	1	5	—	-	-
3. Car, Hotel, ground service reservations	90	25	100	20	1	2.4
4. Airline reservations	90	25	100	20	1	2.4
5. Medical Advice & consultation	-	-	-	30	1	2.4
6. Miscellaneous	?					

Figure 7.24

ACUMENICS

b. Future ATC Concepts

Figure 7.25 lists the major subsystems that determine the aviation communication requirements. Depending on the technology assessment, some or all of the ATC system could be space-based. Fortunately, the increased performance and traffic requirements go together with increased availability of higher technology, particularly space-based technology. To develop aviation communication concepts for the year 2020, the approach employed here is to forecast what the rest of the ATC system is likely to be in order to support the forecast traffic. Only when the navigation, surveillance, and related terrestrial systems have been identified for each scenario can the communication system be developed.

Today, all studies project a major growth in use of telecommunications for a host of office and administrative services, such as office automation, teleconferencing, training and education, electronic mail, remote maintenance and monitoring, use of data collection platforms for monitoring inaccessible facilites, etc. The scope of ground-to-ground voice and data communications can be defined broadly to include these administrative uses more narrowly as illustrated in Figure 7.5 and 7.6. In either case the commercial communication systems available in the year 2020 will be a mix of satellites, optical fibers, various microwave and

Major Subsystems Determining Aviation Communications

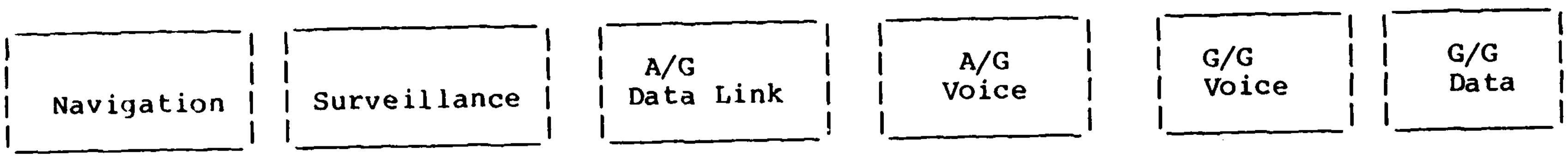


Figure 7.25

other terrestrial systems. A dedicated ground-to-ground communications system for civil aviation is unlikely to be cost-effective. Consequently, the ground-to-ground voice and data communication system is assumed to be essentially commercial in all three scenarios.

In Figure 7.26 three system alternatives are selected to watch the future socio-economic forecasts. The only assumption in Figure 7.26 is how the various ATC systems are likely to be partitioned between space and terrestrial. This then sets up the environment for the aviation communications system.

In the rapid growth scenario, technology will provide for all functions of the CONUS ATC system to be space-based, except for ground-to-ground communications, which will be a mix of space and terrestrial systems.

In the balanced growth scenario several of the functions continue to be terrestrially-based but are aided by satellites to extend coverage and capacity. The satellite-aided concepts would allow for lower reliability spacecraft requirements, less in orbit space capability, etc. In addition, the fully satellite-based system requires multiple beam satellites to make efficient use of spectrum. The satellite-aided concept requires fewer beams, hence it is less technologically advanced. Satellite-aided systems would also be evolutionary; however, they would require

System Alternatives Matched with Future ATC Scenarios

Rapid Growth	Space-Based	Nav.	Surv.	A/G Data	A/G Voice	G/G Voice	G/G Data	Mainly Commercial
	Terrestrial	.	.	.	.			
Balanced Growth	Space-Based	Nav.	Sat. Aided	Sat. Aided	Sat. Aided	G/G Voice	G/G Data	Mainly Commercial
	Terrestrial	.	Surv.	A/G Data	A/G Voice			
Stagflation	Space-Based	Mil. Nav.	.	.	.		G/G Data	Mainly Commercial
	Terrestrial	Nav.	Surv.	A/G Data	A/G Voice	G/G Voice		

Figure 7.26

different rules for frequency assignment between terrestrial and satellite. Finally, in the stagflation scenario, the terrestrial system, as projected in FAA plans for the late nineties, is assumed to change little in configuration. The military navigation system will probably evolve into GPS, but the VOR/DME will continue as the chief civil aid. However, due to the communication industry's current levels of capability, a large fraction of the ground-to-ground communications will be space based, even in this scenario. Potential aviation communication requirements can now be treated further in terms of the three conceptual configurations.

c. Degree of Independence between Surveillance and Navigation

ICAO and FAA have strived to maintain independence between surveillance and navigation systems, so that each can back up the other in case of failures or discrepancies.

Independent surveillance is provided currently by primary radar: its signals and systems are independent of the navaids and communications system and the aircraft avionics. Secondary radar (ATCRBS) relies on an aircraft's transponder and latimeter, so that it is only partially independent in that the same altimeter is part of the aircraft's navigation system; in addition, the aircraft's transponder is essential.

Dependent surveillance is the concept of transferring navigation position data by an aircraft via a communication link to an ATC facility. If used exclusively, navigation system failure would also result in surveillance system breakdown. Redundancy in the navigation system and in the data-link, however, can be added to achieve similar reliabilities as the independent system.

In satellite-aided ATC systems, independence analagous to that of primary radar is not economically feasible, since commonality of the space segment produces significant cost savings. Therefore, criteria will ultimately have to be established in terms of reliability, availability, safety, etc., instead of completely independent or fully dependent. In the system concepts presented here, rough judgments have been made in maintaining the same level of system performance but varying the degree of independence both in terms of signal structure and system hardware.

7.4 Space-Based Navigation, Surveillance and Communication System: Rapid Growth Scenario

The likelihood of a space-based system in the rapid growth scenario is due to the low operations and maintenance costs of such systems due to significantly reduced terrestrial facilities and reduced manpower. In addition, the space-based system will be able to offer higher coverage and improved performance necessary for the huge traffic increase forecast in Figure 7.20. A relevant example for comparision of the cost-effectiveness of space-based systems is NASA's TDRSS (Tracking and Data Relay Satellite System), which is to replace the majority of current ground based NASA tracking stations next year with two operational geosynchronous satellites and two in-orbit spares. This system will cost about one-half the current cost of $200 million per year, and will provide a significant increase in coverage.

a. System Description

In Figure 7.27, the space-based configuration is shown to differentiate the main space, ground, and user segments. The system is composed of a constellation of low-orbit and geosynchronous satellites. A centrally located Satellite Control Center (SCC) collects and processes the surveillance data, and disseminates this data to the various ATC facilities, mainly via commercial communications satellites. Terrestrial links fill in the short distance connections and various commercial telephone networks.

Major Elements of Space Based ATC System

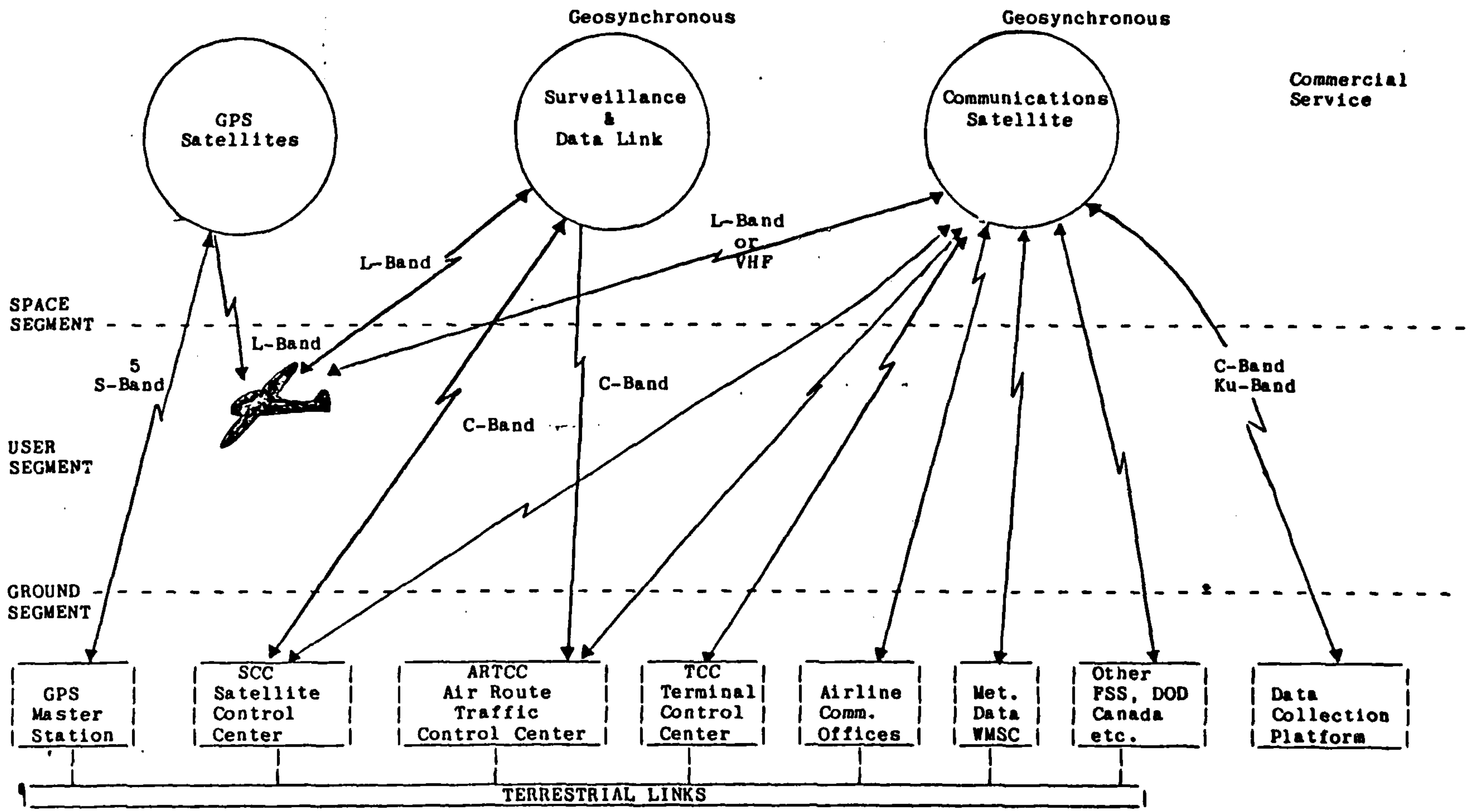


Figure 7.27

Considerable consolidation would become feasible in most of the terrestrial facilities.

The number of enroute Air Route Traffic Control Centers (ARTCC) could be reduced to, at most, 4 or 5. The Flight Service Stations could be significantly reduced by integrating their services into the ARTCC and TCCs. The meterological data could be broadcast, ideally, via the communication's satellite to the aircraft, airlines, and other ATC users. Interfaces with DOD, Canada, etc., would also connect mainly through the satellite communication system.

Consolidation of the terrestrial facilities would produce lighter voice and data requirements per facility, making the cost-effectiveness of the communication satellite network even more effective. The communications satellite could monitor various unattended and automated facilities by means of data collection platforms. The platform data would be communicated to the appropriate terrestrial center--e.g., a centralized maintenance facility.

The basic navigation and surveillance system is GPS based; it is one of several variations analyzed in Elrod et al [4]. For surveillance, position data computed on the aircraft from the relative times-of-arrival (TOA) of ranging signals from 4 satellites is reported via the geosynchronous satellite data link.

In the surveillance process, sufficient time is allocated for both aircraft acquisition and tracking. The acquisition mode involves determining the identity and initial location of new aircraft entering the system. Once an adequate degree of information is obtained, the aircraft is placed into the tracking mode. In this mode, aircraft regularly provide position data to the ground via one of two direct access procedures: polling or time-division-multiple access (TDMA).

With polling, aircraft are discretely interrogated, once each surveillance update cycle, and respond accordingly. An interrogation contains the aircraft address, appropriate ATC messages and it possibly requested for other information. Analysis of the capacity of polling versus TDMA shows polling to be superior [4].

With TDMA, the aircraft, after acquisition, is assigned a time slot in which to respond during each update cycle. Sufficiently accurate knowledge of system time is required on-board the aircraft, which comes from GPS. Either TDMA or polling associated with surveillance is a data link. This data link can be employed for non-navigation or non-surveillance air-to-ground communications, but the system would have to significantly change to handle the higher capacity. Consequently, the communications satellite handles all the other data and voice communications.

Navigation and surveillance data are from the source 4 GPS satellites. Independence between the two functions is obtained by (a) employing separate GPS receivers and, (b) verifying aircraft derived range by comparing it with an independently derived range measurement based on the TOA of an aircraft surveillance reply.

Figure 7-28 illustrates the satellite constellation for the proposed concept. The GPS satellites are in 12-hour, 63^{o} inclined orbits. The Surveillance Data Link satellites (SD) are in geosynchronous orbit. In the year 2020, a single satellite could be produced to serve the full capacity requirements of CONUS surveillance. A second satellite would be included as an in-orbit backup. A second operating satellite may be added to improve GPS accuracy in some areas [4].

The Surveillance SD Satellite communicates with aircraft on L-band and to the SCC on C-band. The SD Satellites could also be used to assist the aircraft navigation receiver in selection of the 4 GPS satellites to use, calibration corrections, and aiding acquisition of the GPS receiver. Many of these concepts are being studied for reduction of GPS user costs. To maintain the high level of system integrity, the system concept includes Calibration stations deployed around CONUS for providing correction data (e.g. propagation delays) to maintain both navigation and surveillance accuracy.

Space-Based Surveillance/Data

System Constellation for CONUS
(Voice and Communications Satellites are separate)

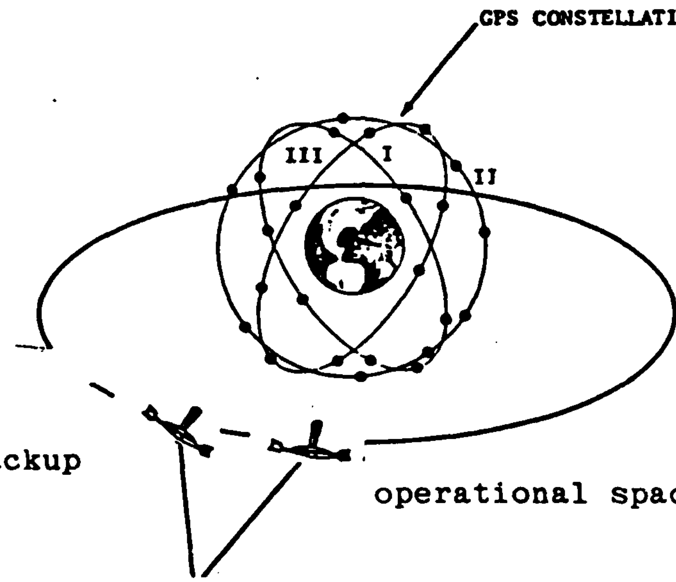


Figure 7.28

The majority of air-ground and ground-fraud communication traffic is-on the geosynchronous commercial satellites. The aircraft-to-satellite link will be either at L-band or VHF, while the satellite-to-ground link will be at either C-band or Ku-band, as assigned for commercial communications. As far as the system concept is concerned, the air-to-ground communications may be on one satellite and the ground-to-ground communications on another.

The space, user, and ground segments are further discussed below, in terms of the communications system. The related surveillance navigation system is further described in Appendix A.

b. Communications Subsystem for the Space Based Concept

In this section, the communications subsystem for the space-based concept is further developed. To do this, it is useful to identify the following major parts of the subsystem:

1. Air-ground communications in which there are large numbers of users:

   - aircraft 700,000
   - ground
     - terminals 2,000-5,000
     - airlines
     - airports
     - TCC

Capacity of this part is principally set by air-to-ground voice frequency.

2. Ground-to-ground communications between the major facilities such as the SCC, ARTCC, National Flow Control Center, Centralized FSS's, National Weather Service, Central Airline Offices and TCC's should, at most, add up to 20-100. Capacity of this link is determined mainly by data requirements.

3. Ground-to-ground communications within major facilities--basically the large and medium hub airports. Capacity here is small for current services, but could grow significantly with new uses of communications.

4. Ground-to-ground communications from large numbers of data collection platforms to a few centralized facilities. This would be the basic configuration for the Remote Monitoring and Maintenance System (RMMS) for the various smaller terrestrial facilities.

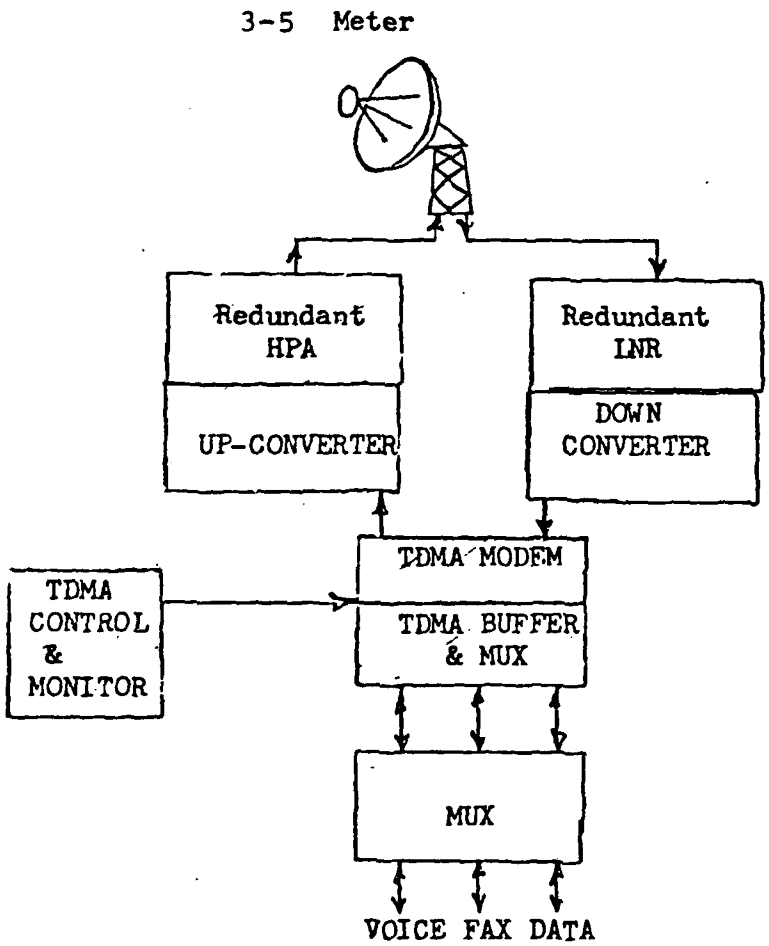
Except for intefacility communications of category (3), all others above could be integrated into a single communications satellite. However, due to the special nature of the air-ground communications, which require low-cost user terminals, it is desirable to separate the space segment into two types of spacecraft. For the air-ground communications, a multi-beam satellite with high power and voice links is selected. For the ground-to-ground communications of category (2), a data communications domestic satellite service is assumed. This same service, in addition, could be used to sequentially collect data from all the DCP's and send it to a central facility.

In the year 2020, domestic satellite traffic will be served, at least, in the C-band and Ku-band and the 20/30 GH_3 band. Due to the propagation problems at the higher bands, the satellite service for category (2) should be mainly at C-band. Ku-band should be feasible in most areas except for regions with very heavy rainfalls. Technology forecasts indicate that the most effective access schemes will be TDMA. Typical terminals will have 3-5 meter antennas, and the satellites will employ satellite beam switching, making them cost-effective by keeping the major ATC facilities in as few beams as possible. A block diagram of a TDMA terminal is shown in Figure 7.29. There are many projections of the configurations of future domestic satellites for wideband tracking -- the best for category (2) service would be simply selected on the basis of commercial offering.

c. Air-Ground Communications Satellite Concept

In the case of a system concept for air-ground communications, further specificity is necessary. In Figure 7.30, the proposed multi-beam satellite for air-ground communications is shown. This satellite concept is based on a study by General Electric Company sponsored by NASA. The concept is developed in significant detail in Reference [13], for application to future land-mobile communications; it is adapted here with few changes. However, the proposed system capacity and concepts of modulation,

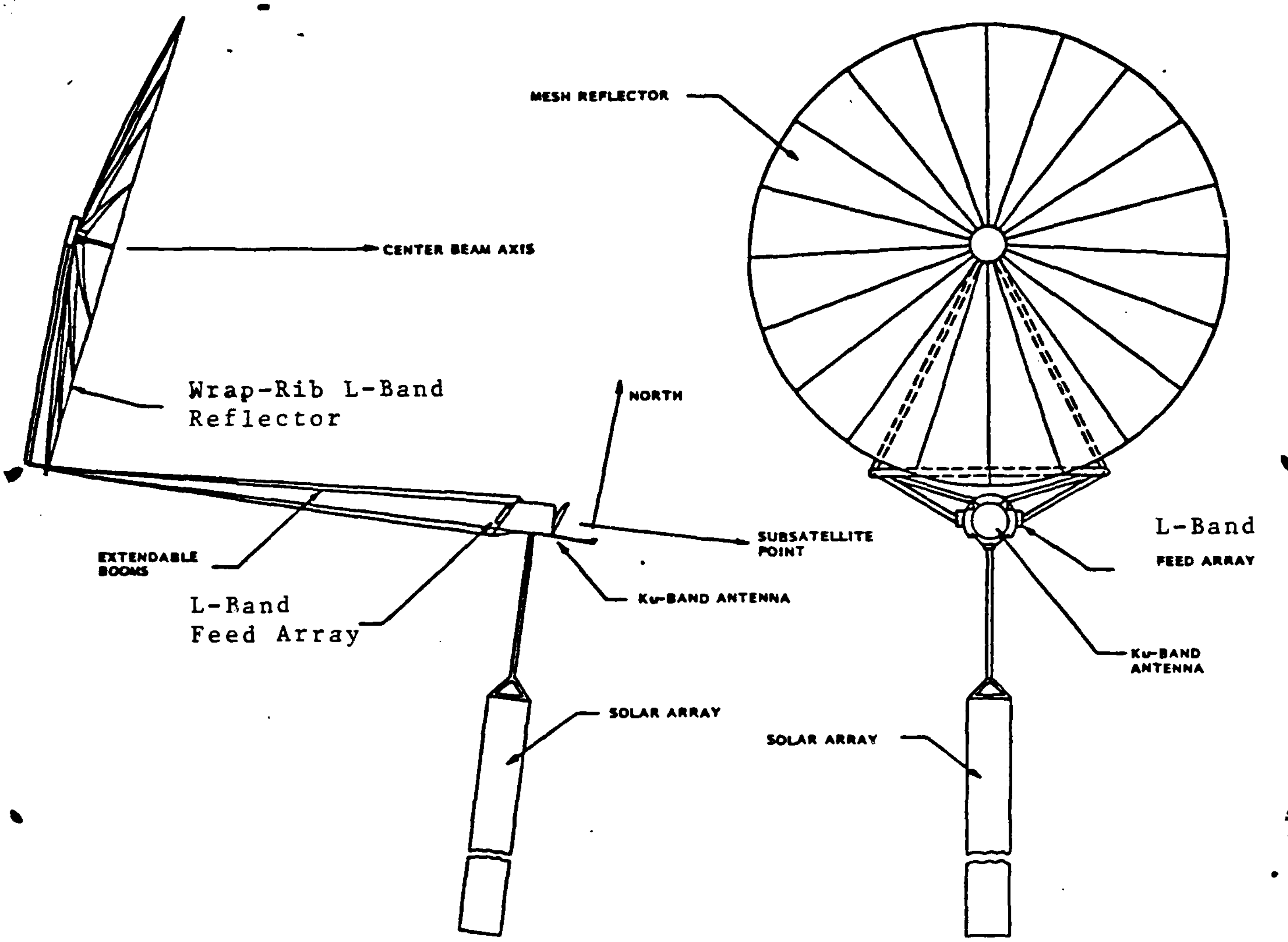
A Typical Terminal for Ground-to-Ground
Communications via Satellite



TDMA - Time Division Multiple Access
LNR - Low Noise Receiver
HPA - High Power Amplifier
MUX - Multiplexer

Figure 7-29

Spacecraft Design Applicable
for Air/Ground Aviation Communications



G.E. study of reference [13]

Figure 7.30

multiplexing, performance, etc., may be modified with little impact on the viability and capability of the basic issue of air-ground communications via satellite in the rapid growth scenario.

The terrestrial mobile telephone market is growing rapidly, and the impact of this mass market can be used in designing the air-ground communications to have similar narrow-bound FM channels. These channels are allocated 30 KH_3, each with IF bandwidth of 25KH_3 and a voice message bandwidth of 300H_3-3KH_3. Assuming a 10MH_3 bandwidth from aircraft to satellite, and a 10MH_3 from satellite to aircraft, permits a total of 333 duplex voice channels. A single beam satellite would not be sufficient to serve the 100,000 or more peak CONUS IAC traffic for the rapid growth scenario. A 69 beam satellite with the pattern shown in Figure 7.31 would be able to increase the frequency reuse by a factor of 23.

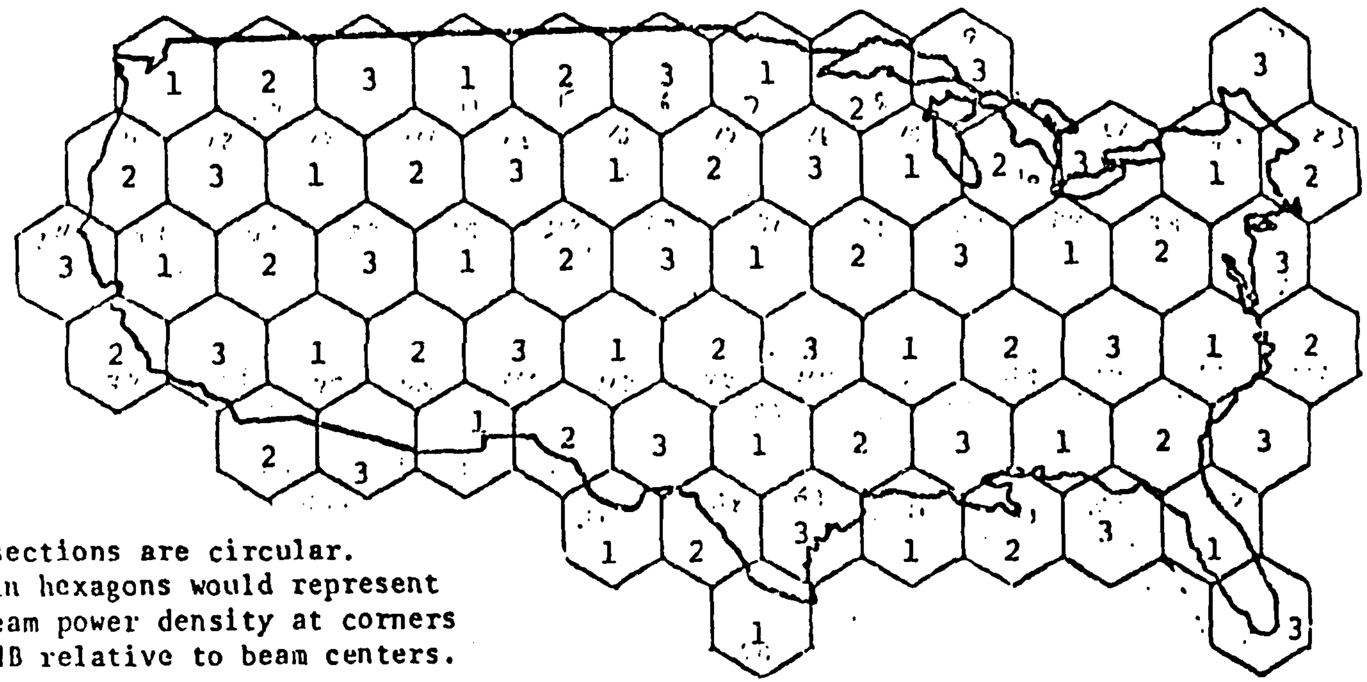
The 333 channels within the allocated frequency band are divided into 3 sets as follows:

Set 1: 1, 4, 7, 10, 13 ... 331

Set 2: 2, 5, 8, 11, 14 ... 332

Set 3: 3, 6, 9, 12, 15 ... 333

Satellite "Cells" Serving Contiguous States



Actual beam cross sections are circular. Circles inscribed in hexagons would represent 0.5° beamwidth. Beam power density at corners of hexagons is -4 dB relative to beam centers.

Numbers in hexagons refer to channel sets of frequency reuse pattern.

Figure 7.3

The numbers within the cells of Figure 7.31 represent the channel set in the beam. Note that no adjacent beams have the same set of channels in order to minimize co-channel interference. Each beam or footprint of the satellite has 111 channels, each of transmission and reception.

Every aircraft in the system is capable of operating on all channels in all sets of frequencies. Within each set of frequencies there is at least one "calling" (aircraft-to-satellite) and one "control" (satellite-to-aircraft) channel. When an aircraft's unit is activated, it automatically searches and locks onto a control channel. When the pilot dials a number, the request is transmitted over the calling channel. The calling channel signal goes to the satellite control station, which assigns a talking channel pair to the aircraft via the control channel, and connects the call to the called party through the public terrestrial network.

The transponder concept assumes a satellite switch which has 69 L-band inputs and 69 Ku-band outputs as shown in Figure 7.32. A 69x69 satellite switch is more advanced than today's communications satellites. Currently, the TDRSS/Advanced Westar is the only COMSAT which will have a 4x4 beam switch. However, on-board switching, as indicated in Figure 7.32, will be available by the early 1990's, since it significantly reduces earth station switching requirements and it can eliminate double hops in aircraft-to-aircraft communication

Basic Satellite Switching Configuration

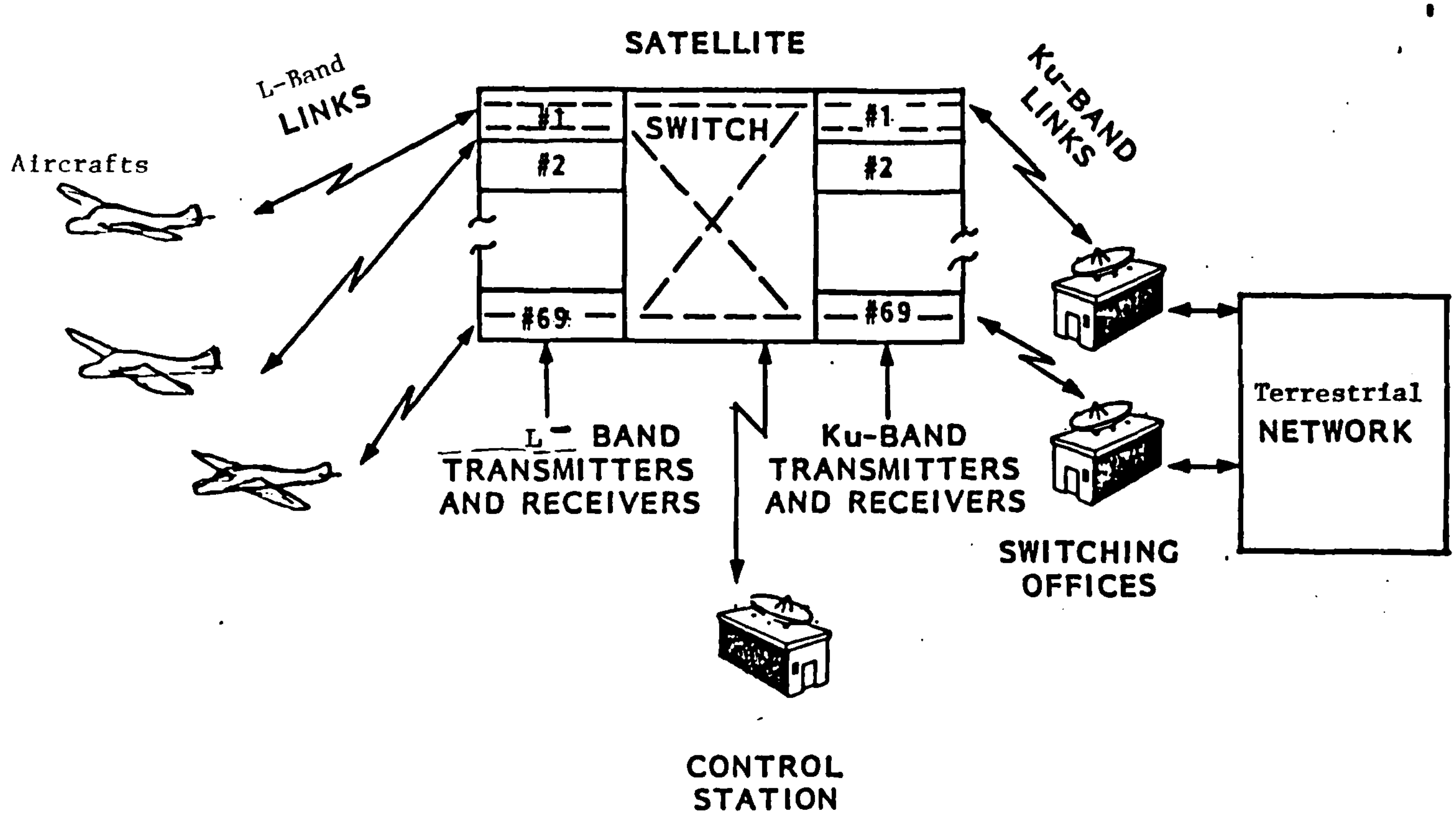


Figure 7.32

Figure 7.33 is a more detailed description of the satellite transponder. The figure shows each beam's FDM channels as being down-converted (D/C) to an IF demultiplexed and then switched to the addressed destination beam. The switch outputs are appropriately processed, multiplexed and up-converted (U/C) before retransmission. The processing required ahead of the switching can range from down-converting to a convenient IF and separating the channels with filters, to demodulation of each channel followed by analog-to-digital conversion, digital switching, digital-to-analog conversion, remodulation and multiplexing for transmission. Such an advance switch eliminates the switching required at the earth stations and may be a significant cost savings in total operation and maintenance costs.

Reference [13] considers a non-switching transponder satellite design. It would be a lower order of technology, and it will have double hops in the aircraft-to-aircraft links. Other work has been reported by COMSAT for a TDMA concept instead of the FDMA approach presented here. However, that concept was developed for a lower capacity -- although it could also be extended [14].

System Capacity

Capacity is a function of the grade of service and the number of channels. The "Erlang" is the international dimensionless unit

Block Diagram of Transponder

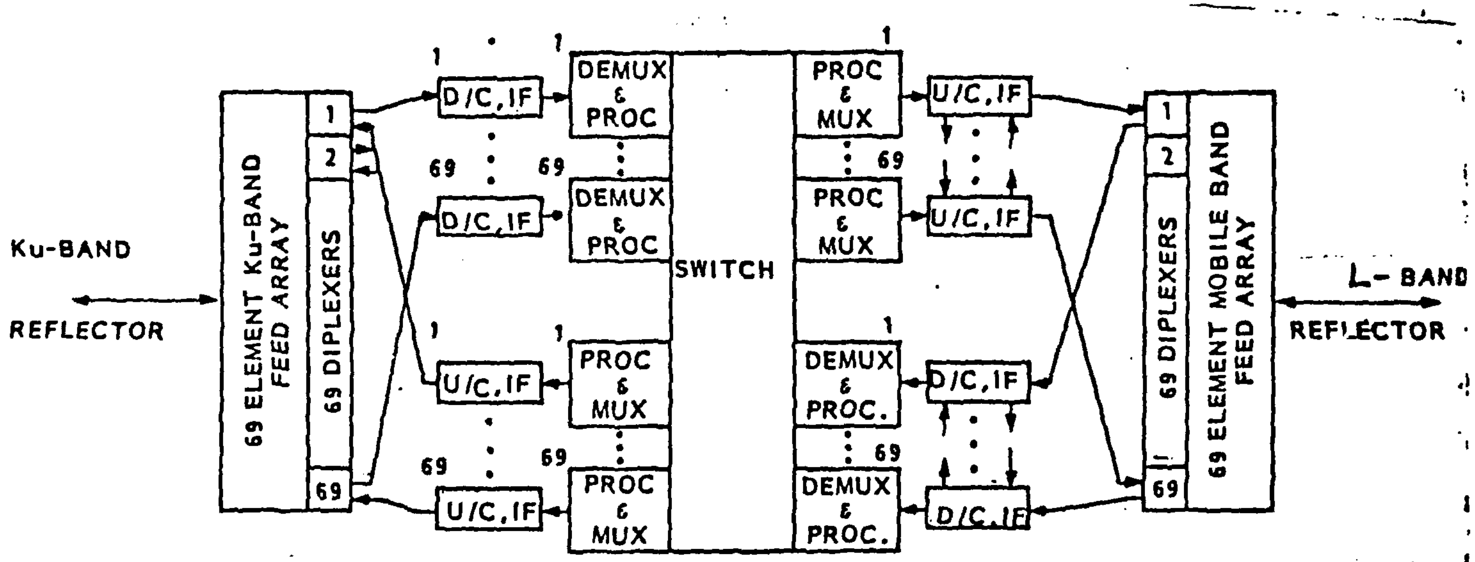


Figure 7.33

of traffic intensity defined as the intensity in a traffic pattern continuously occupied. The capacity of a single footprint with 108 channels, and a service grade probability of 2 percent blocking comes to 95 Erlangs given by:

$$B(C,A) = \frac{A^C/C!}{\sum_{k=0}^{C} A^k/k!}$$

where

B(C,A) = probability of a call being blocked

A = traffic intensity to be served, in Erlangs

C = total number of channels

It is further assumed in this deviation that calls not immediately satisfied are cleared and do not reappear during the period under consideration.

Assuming an average duration of 120 seconds per call, 95 calls can be satisfied in one foot print every 120 seconds, or 2850 calls maximum per hour. The total instantaneous capacity of the system is 196,659 for all 69 beams. However, the traffic will not be homogenously distributed so that the peak capacity would be lower. Further traffic distribution analysis would be required to get a more accurate system specification.

Spacecraft Technical and Weight Estimates

Reference [13] describes the satellite power requirements at the L-band to be 48 kilowatts, with an L-band antenna of 70 feet in diameter. Figure 7.34 gives the spacecraft technical characteristics and Figure 7.35 gives the weight estimates. A conservative cost of such a satellite system in the late 1980's is $58 million/year not including user equipment costs. In the year 2020, both the high-power and the satellite switch will be significantly lower in cost if the assessment of technology growth assumed is realized. This concept is based on compatibility with current terrestrial mobile terminals. If for ATC the mobile terminals have higher radiated power and lower receiver nose figures, the satellite power and cost can be further reduced. Analysis of these alternatives are not within the scope of this report.

Avionics

The avionics will be a straightforward FDM/FM unit as shown in Figure 7.36, together with a voice input. The system is likely to include data transmission. The typical cost of such an avionics unit should be no more than today's VHF two-way voice unit, which is in the range of $1,000 to $2,000. However, the quality of the proposed FM will be significantly superior to the current AM system. The characteristics of the avionics for this system concept are the following:

Spacecraft Technical Data

Payload	
Forward Transponder	69 Parallel Redundant Units Ku-band to Ku-band Frequency Conversion
Return Transponders	69 Parallel Redundant Units L-band to Ku-band Frequency Conversion
Transponder Type	Double Frequency Conversion
EIRP per Voice Channel	45.2 dBW
EIRP per Beam (Transponder)	65.7 dBW
Number of Beams	69
Voice Channels per Beam	111
Antenna Gain	44 d Bic - At Corner of Cell
RF Radiated Power per Beam	21.7 dBW, 148 Watts
G/T	16.4 dB/^{o}K
Bandwidth per Beam	10 MHz each way
L-band Antenna	
Type	Offset Parabolic Reflector Multibeam, F/D =
Effective Aperture	75 ft Diameter
Spacecraft	
Size	15 ft L, 9 ft H, 15 ft W
Weight	5515 lbs
Electrical Power S/S	41 kW Array Power 5130 ft^2 Solar Array
Thermal S/S	1200 ft^2 Radiating Surface

Figure 7.34

Spacecraft Weight Estimates (Pound)

		L-band
Mobile Band Antenna		
-	Reflector	250
-	Supports	125
-	Feed Array	160
Transponder		
-	Mobile Band Diplexers	35
-	Mobile Band Power Amplifiers	140
-	IF Components	100
-	Ku-band Power Amplifiers	15
-	Ku-band Diplexers	15
-	Cabling	50
Ku-band Antenna		
-	Reflector (solid)	35
-	Supports	10
-	Feed Array	5
Spacecraft Structure		700
Propulsion		1200
Electrical Power Subsystem		
-	Solar Array	2100
-	Electronics	75
-	Battery	450
TT&C		50
	Estimated Weight-pounds	5515

Figure 7.35

Air-Ground Voice Avionics

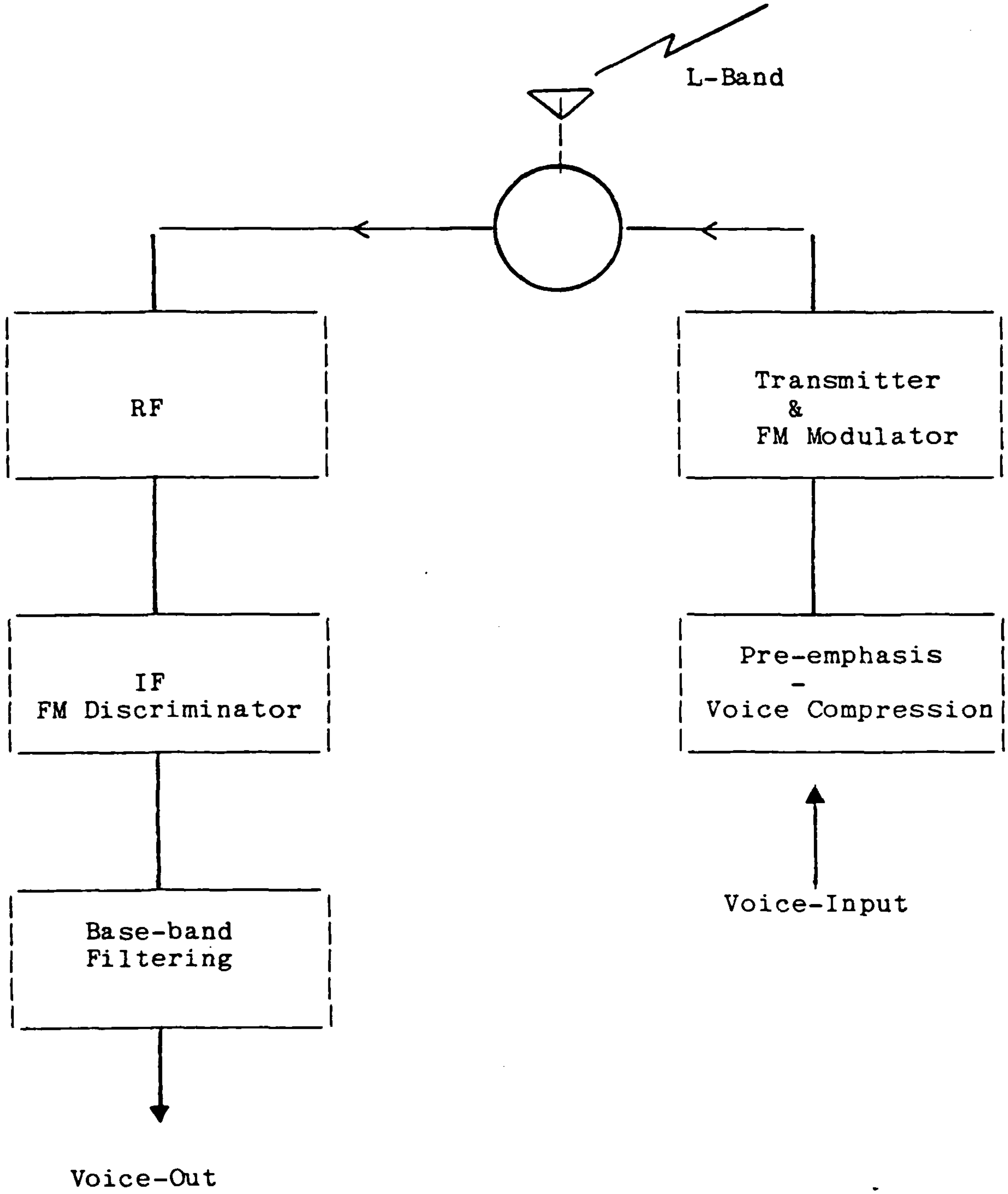


Figure 7.36

Signal quality:	50 db subjective S/N minimum
Coverage:	Whole of CONUS to all altitude levels where aircraft can safely traverse
Modulation:	Frequency modulation with $12KH_3$ deviation
Channels:	30KHG bandwidth at L-band ($1500-1600MH_3$)
Receiver Nose Temperature:	$546^{o}k$
Transmitter Power:	10 watts
Aircraft Antenna:	Omni-directional gain of 4dBIC

The 50db S/N ratio is obtained by use of emphasis circuits and compounding. The system concept could easily be modified to digital voice employing Continuously Variable Slope Delta Modulation (CVSD).

d. Conclusions

The high traffic growth forecast for the rapid growth scenario will bring about major changes to the current surveillance, navigation and communication system. Fortunately, in this scenario, there will also be a major growth in technology and telecommunications. Consequently, the availability of a completely space-based system for ATC functions is forecast for the year 2020. The cost-effectiveness and increased coverage of the space based system will be the reason for its implementation. In such an environment, aviation communication is separated into a trunking communication service between the major ATC ground stations and air-ground communications.

The trunking service between the relatively few facilities, which are still likely to be distributed over CONUS, is assumed to be served by commercial domestic communications satellites. However, for the air-ground voice communications, a separate satellite concept is presented. Evolution of such a system is expected to develop from MARISAT, an Oceanic Satellite ATC system such as Aerosat, and other concepts currently being developed for a land-mobile telephone system.

7.5 Hybrid Terrestrial and Space Based ATC System

In the balanced growth scenario the traffic growth is gradual, with sufficient time available to expand gradually the performance and capacity of the ATC system. Central to the balanced growth scenario will be the two systems: (1) GPS for navigation and (2) for surveillance, ground-based DABS with satellites aiding in the expansion of coverage and transition to a future space-based surveillance system.

a. System Concept

A satellite based navigation will remove the 1000 VOR/TACAN/DME sites, communication requirements to and from those sites, and the frequency allocated to VOR and TACAN/DME. Elimination of VOR together with ILS would also impact upon the Flight Service Station network and VHF voice. The advantage of L-band for satellites versus VHF would indicate the desirability of switching to L-band for air-ground voice. However, as the technology advances are less than in the rapid growth scenario, the air-ground voice would be mainly terrestrially-based with a less complex satellite providing coverage in areas outside the terrestrially-based stations.

The advances in computational capability, sophisticated displays, and further automation will allow centralization of the ARTCC's and Flight Services, Flow Control, FAA maintenance and operations. Weather information would be centrally processed and disseminated by the DABS data link and the air-ground voice communications to aircraft.

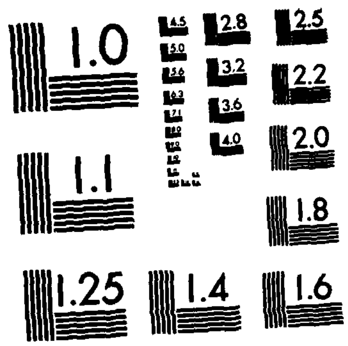
AD-A126 015 SOCIOECONOMIC IMPACT ASSESSMENT: COMMUNICATIONS INNUSTRY PHASE III TECHNO..(U) ACUMENICS RESEARCH AND TECHNOLOGY INC BETHESDA MD 02 FEB 79 FAA-APO-81-11-4 7/7

UNCLASSIFIED DOT-FA78WAI-932 F/G 17/2.1 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

The terrestrial components of the DABS surveillance system would be deployed in the major hub areas mainly, at terminals. Satellite surveillance compatible with terrestrial DABS would fill in the remaining coverage areas. To accomplish this the satellite would have to be powerful enough to satisfy the aircraft's DABS transponder receiver, and the satellite receiver would have to be able to pick up the DABS transponder transmissions.

Ground-to-ground communications with space-based GPS and satellite-aided DABS would require communication links from DABS facilities to the ARTCC's and TCC's, in addition to the other centers. The monitoring of terrestrial facilities for reducing costs of operations, maintenance, certification, etc. would involve significant use of data collection platforms, as in the space-based scenario, where data is collected by a satellite transponder. The growth of these platforms for many other monitoring purposes, such as environmental monitors, weather data sensors, ocean bouys, inaccessible military facilities, etc. would result in a variety of commercial offerings. The value of data collection platforms in this scenario is even more important than in the space-based scenario due to its cost-effectiveness. Figure 7.37 describes the major elements of the Hybrid ATC System.

In the 1990's the ARTCC and FSS are expected to consolidate into a single facility. By the year 2020, further centralization of the 20 or so CONUS ARTCC/FSS into, at most, 2-5 facilities

Major Elements of Hybrid ATC System

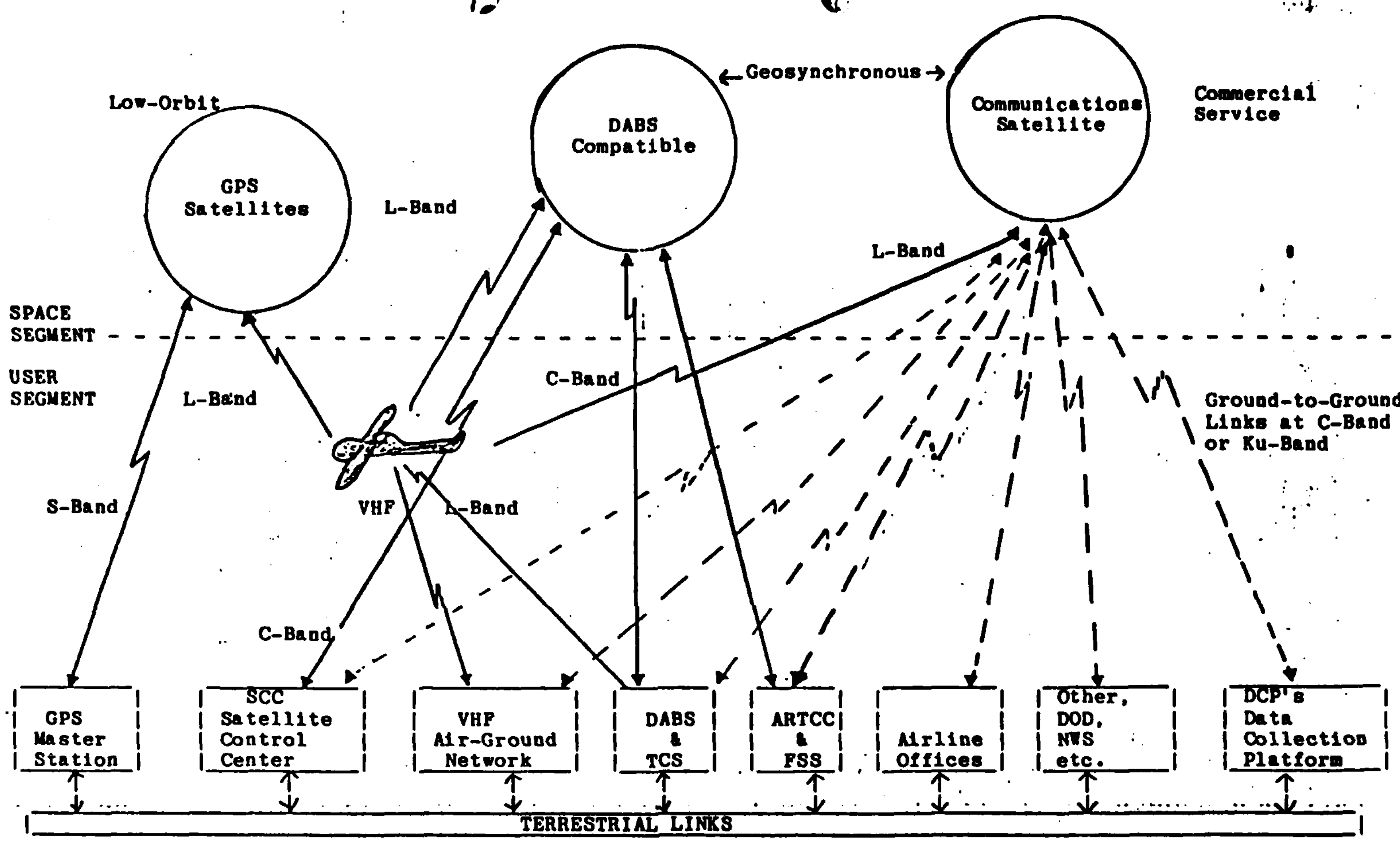


Figure 7.37

would result in longer distance communication links between DABS, TCC's and other terminal facilities to the ARTCC/FSS facilities. Commercial satellite services are likely to be the most cost-effective for ground-to-ground data and voice communications. However, the system would be optimized to cost-effectively integrate terrestrial tails from the satellite earth terminals. Availability and reliability would be maintained by appropriate space-segment back up and selected narrow-band terrestrial links.

By the year 2020, the Domestic Satellite Communications Carriers will have their satellites either linked by intersatellite links or contained on the same large platform with other carriers. This will give a significant increase in space-segment availability for all users of satellite communications.

The consolidation of ARTCCs with FSS will result in increasing internal communication systems within facilities. Fiber optics and efficient data and voice coding methods will find widespread use. Similarly, communications requirements between facilities in the terminal area will increase support, thereby increasing automation in maintenance, monitoring, controller workload, etc.

b. Surveillance

The development of the surveillance system concept is beyond the scope of this task. However, it is useful to define the basic concept implied in the satellite-sided DABS. The need to provide increased coverage within a fixed budget will provide incentives to adopt a satellite system by 2020. An objective of the satellite system is to reduce the number of terrestrial DABS sites from 300 to about 50-100 planned sites. These DABS sites would be located in high density areas. Satellite surveillance would be provided in less dense geographic areas. The surveillance and data link SD satellite will have fewer beams and lower capacity than the full CONUS coverage system described in Section 4.

In areas where DABS coverage is not available, position information for surveillance would be based on GPS. The coordinates of the terrestrial and space systems would be calibrated (4-8). In addition, the SD satellite would be used to connect all DABS stations to the ARTCCs and TCCs. The data link formats would be the same for both the terrestrial and space links.

At present, all terrestrial surveillance and data link system concepts have been planned and are summarized in Appendix A. As noted above, this system includes 300 DABS sites which will be connected with the ARTCCs, TCCs, etc. The terrestrial system requires more data communications capacity than the satellite-aided DABS. However, this mixed system may then be considered as a transition

state betwen the terrestrial system and an all space-based system. As such, the system would be likely under the Balanced Growth Scenario.

c. Communications Subsystem for the Hybrid/Terrestrial Concept

The communications subsystem, as noted in the space-based concept, includes the following major elements:

1. Air-ground communications;
2. Ground-ground communications between major centers and terminals;
3. Ground-ground communications within facilities and major and medium hubs;
4. Communications to data collection platforms (DCP's).

Air-ground voice communication will use primarily VHF. The use of VHF will embody increased use of digital voice coding and improved efficiency of spectrum use. However, high performance users such as air carriers will be offered satellite communications by commercial carriers. The satellite-to-aircraft links will be at L-band requiring a separate L-band RF section for those users. Such a system would require a less complex satellite than described in the Rapid Growth scenario. The likelihood of two air-ground communication systems existing is reasonable since the economic returns from improved communications to the air-carrier and other high performance users will be significant. Another basis for this forecast is the assumption that oceanic ATC communications via satellite will be operational by the year 2020, so for many of the carriers there may be no additional investment required.

The terrestrial VHF towers could also be connected via satellite to an FAA central facility. There are 3 subsystems for air-ground communications in this scenario:

- ATC data-link on DABS;
- L-band/KU-band satellite for high performance users;
- Terrestrial based VHF, with satellite link-up towers.

The satellite communication subsystem is given in the next section.

Ground-to-ground communications between major terminals will be provided as both commerical carrier satellite and terrestrial links. Standardized configurations would exist for major hubs, medium hubs, and small terminals. Intrafacility communications is a highly dynamic and innovative industry and maximum opportunity should be taken in reducing costs and performance of facility operations. As such, communications will be expanded with transmission by fiber optics.

DCP's are basically composed of a sensor and a transmitter. The satellite picks up the data from the sensors and transmits it to central locations. Today non-military DCP's use is in global environment and earth observations, with data being relayed by the weather and land satellites (15). These platforms transmit at 400 MH with data rates up to 5 KBPS at 3-10 hr. intervals. DCP transmitter power is less than 5 watts. Monitoring for maintenance and certification could be easily satisfied within an order of magnitude of these specifications. With use of geosynchronous satellites, continuous monitoring of critical facilities could be maintained. A block diagram of a typical DCP is shown in Figure 7.38.

Typical Data Collection Platform
for Monitoring ATC Facilities

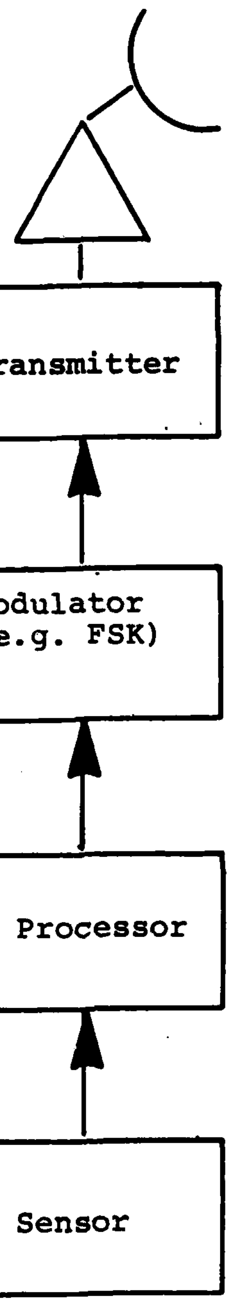


Figure 7.38

Typical Data Collection Platform

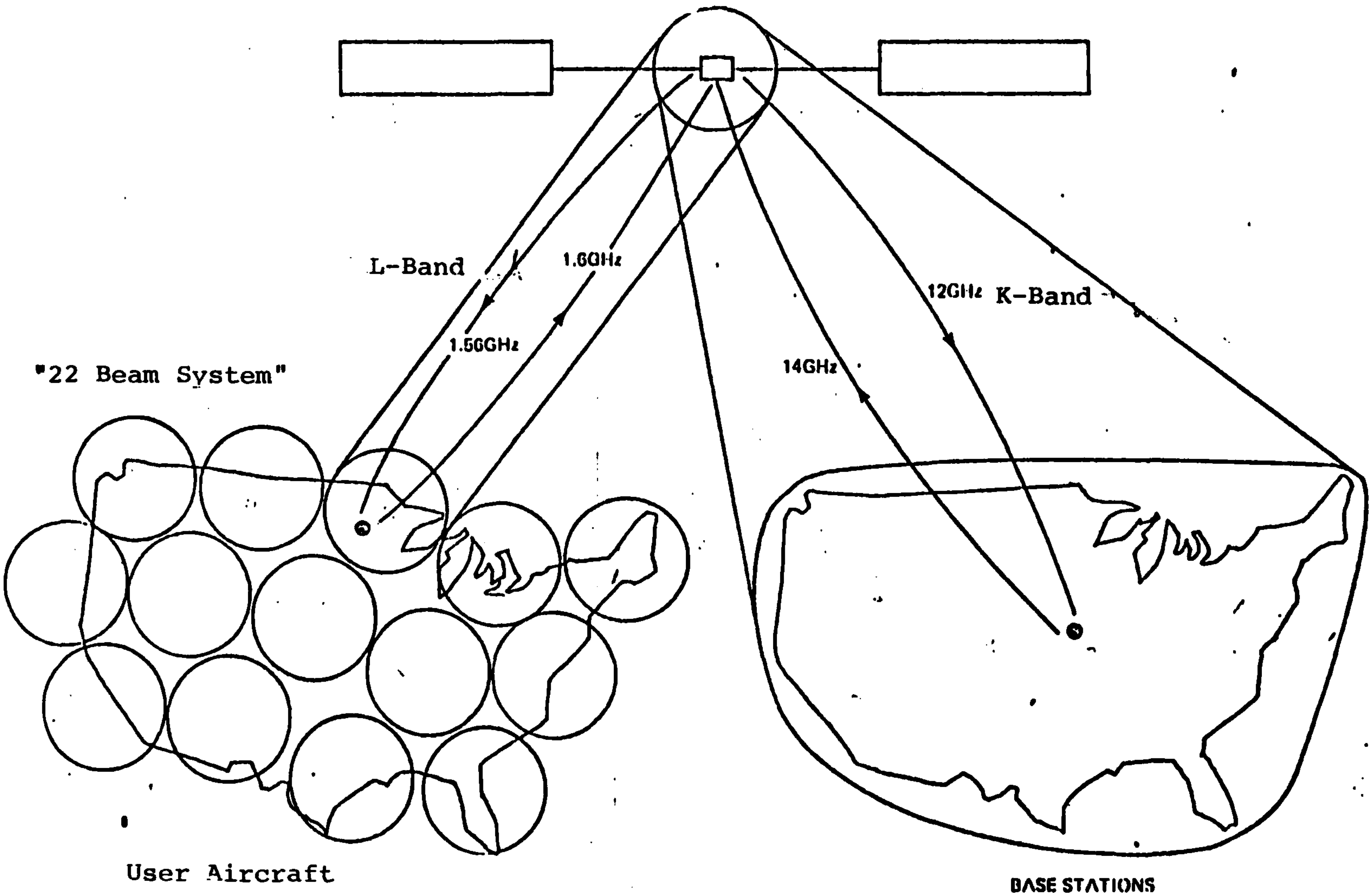
d. Satellite Subsystems for Air-Ground Communications

This section briefly describes the air-ground satellite communication for air-carrier and other high performance users and the use of a satellite for linking the terrestrial VHF towers. An air-ground communication satellite concept derived from a COMSAT study of 1978 (14) is shown in Figure 7.39. This satellite concept is less advanced than the design described in the Rapid Growth scenario. However, the design represents the current state-of-the-art and is being proposed as a new initiative for COMSAT in the 1980's. In this COMSAT design, a 22 beam satellite is proposed, employing TDMA with digital modulation, which could use CPSK, DPSK or FSK. The voice processor has a rate of 16 kbps using CVSD (continuously variable slope delta-modulation) or could use adaptive PCM. The resulting voice intelligibility should exceed 95% with a bit error rate of 10^{-4}.

A sketch of the satellite transponder with onboard switching is shown in Figure 7.40. The corresponding block diagram of the avionics is shown in Figure 7.41 and the user base stations is shown in Figure 7.42. The costs of the avionics and the base stations were estimated at $3,000 and $55,000 respectively in the COMSAT study. Each user group would have its own base station. The cost estimate of the 22-beam satellite system was $24 million annually. Technology forecasts indicate significant reduction in these costs by the year 2020. The economics of scale and a 40-year learning curve should result in similiar user costs for a "satelliteradio" or "terrestrial VHF-radio".

The user cost estimates are given in Figure 7.43, while the satellite mass-budget is given in Figure 7.44. The satellite cost

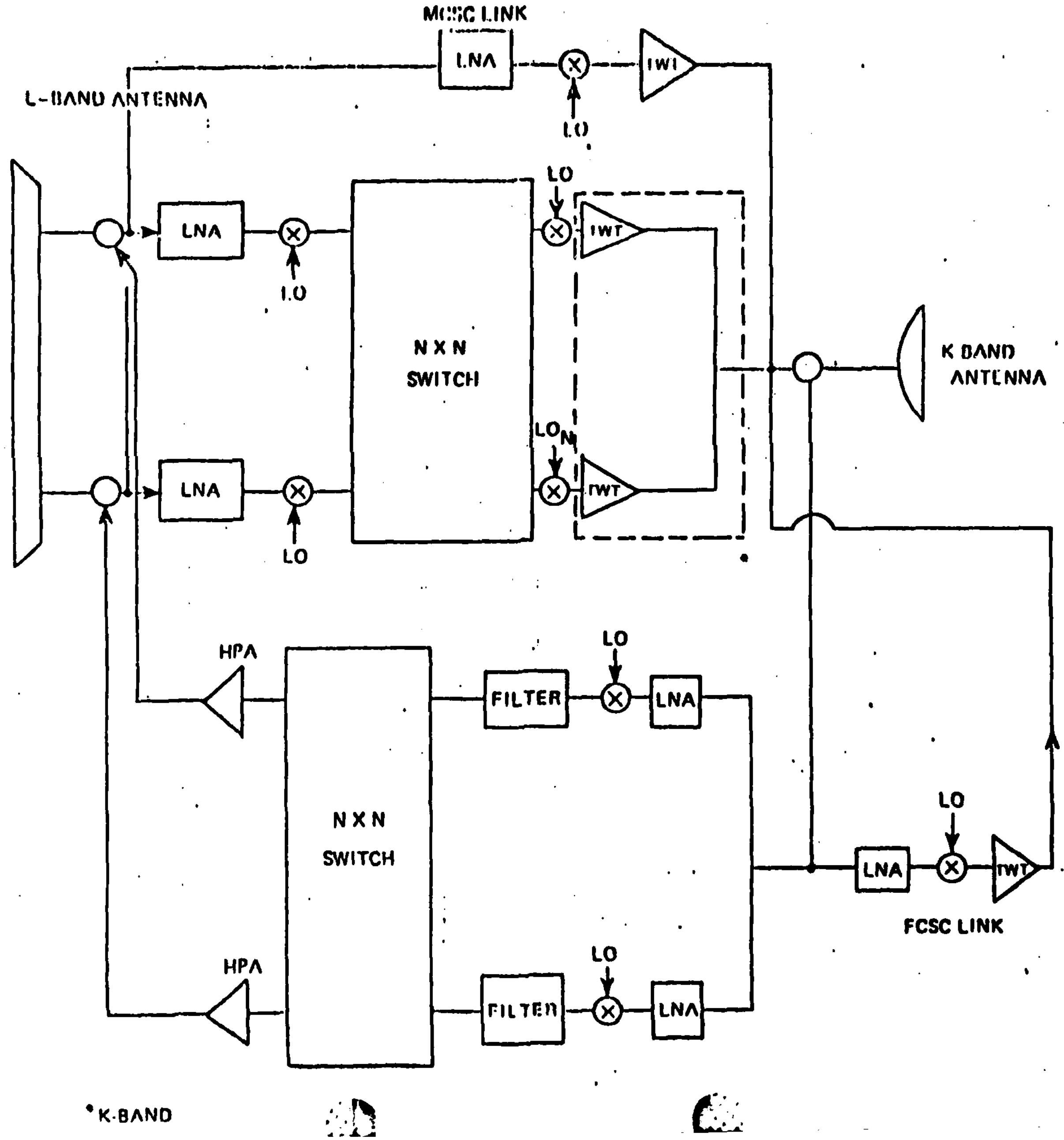
Air Ground Satellite Communications Concept
Derived from Preliminary Study by COMSAT [14]



CONCEPT I

Figure .39

Satellite Switched Transponder
Derived from Preliminary Study by COMSAT [14]



Avionics Block Diagram
Derived from Preliminary Study by COMSAT [14]

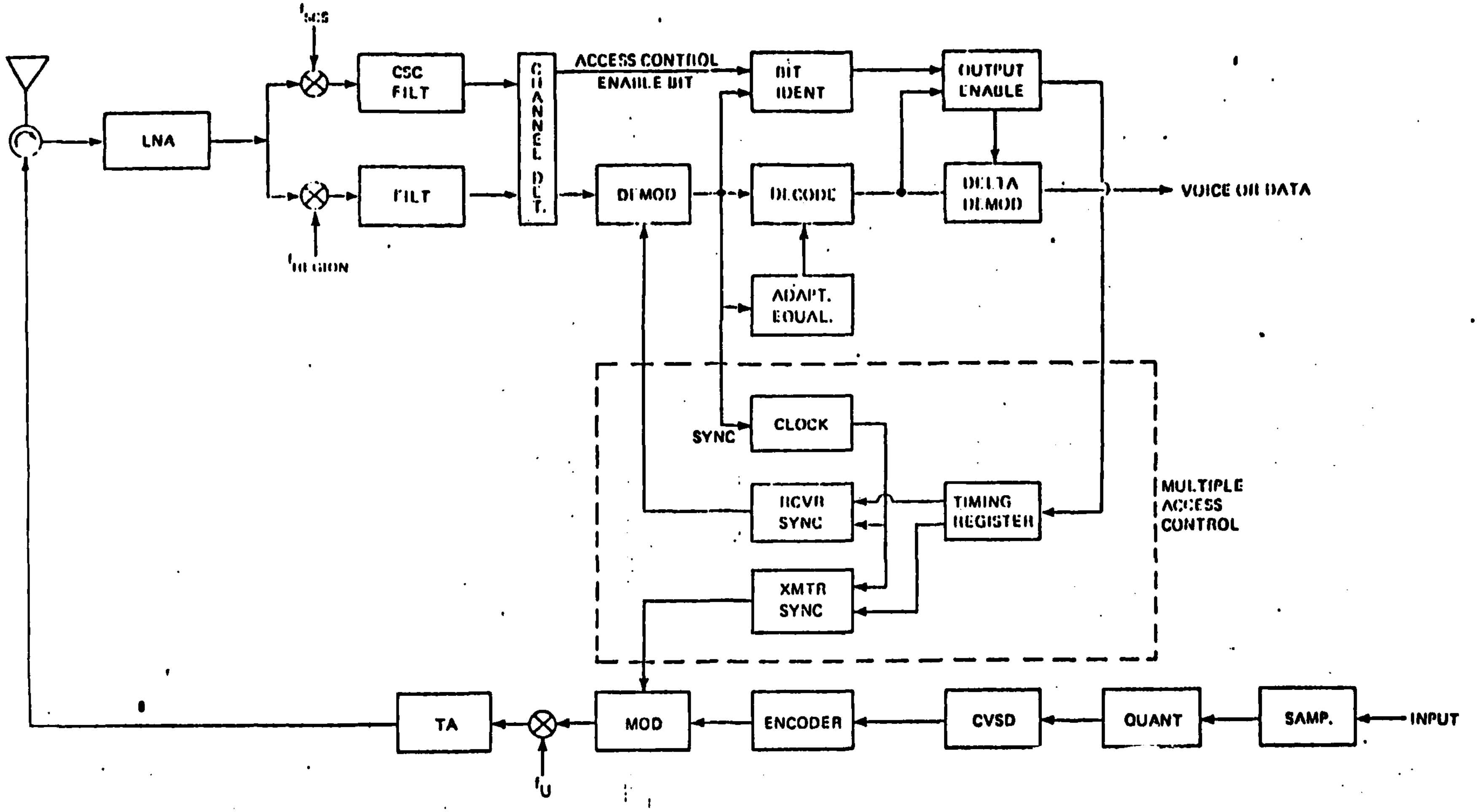
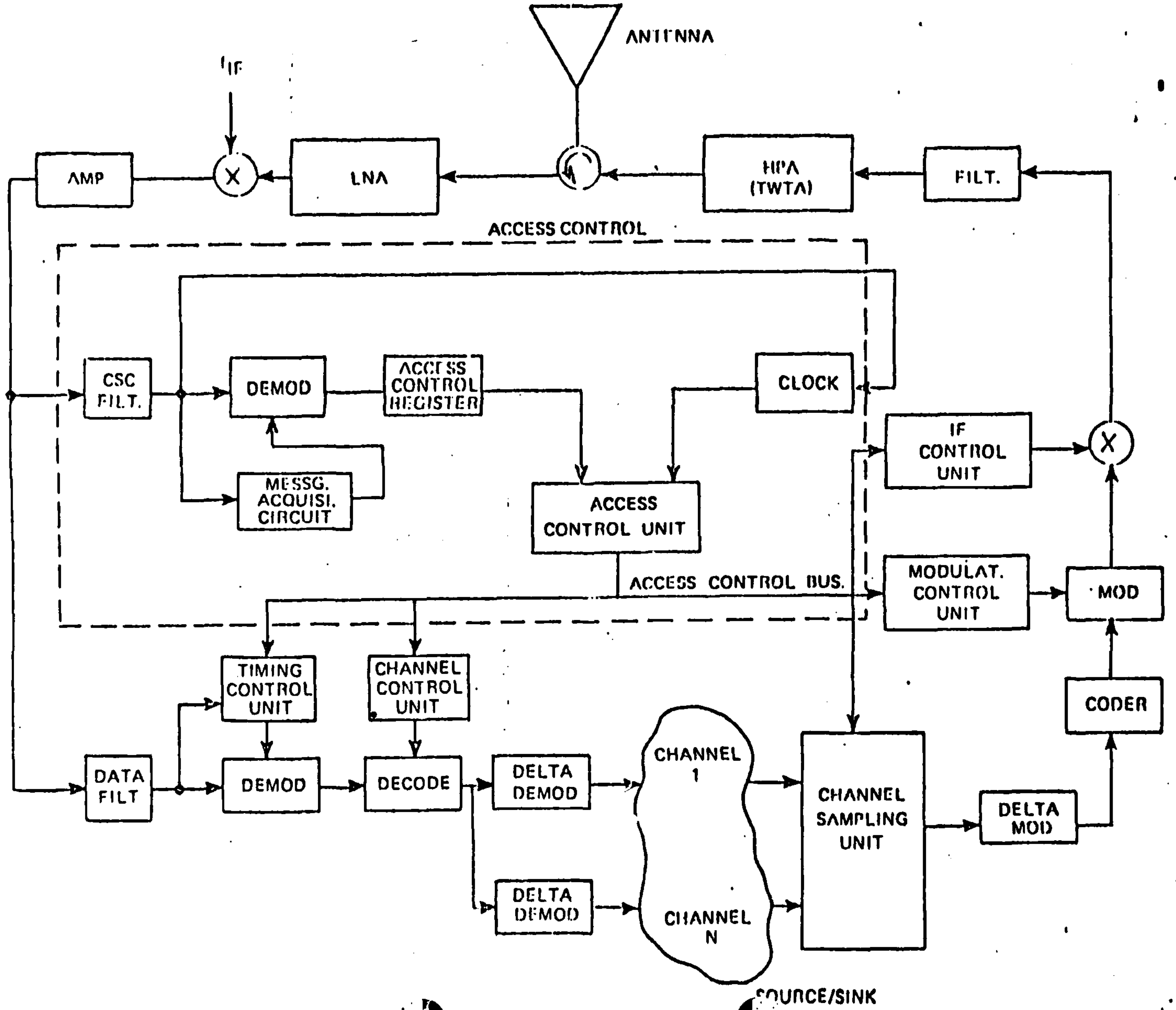


Figure 41

Base Station Block Diagram
Derived from Preliminary Study by COMSAT [14]



User Cost Estimates for Avionics **

COMPONENT	MOBILE TERMINAL DESCRIPTION	MOBILE TERMINAL COST	FIXED BASE STATION DESCRIPTION	FIXED BASE STATION COST
HPA	20 Watt Transistorized power amplifier*	$400	20 Watt TWTA	$2,000
LNA	400° transistorized amplifier*	$100	Uncooled 1000° GaAs FET Amplifier	$5,000
MODEM	PSK 1 Mbps modem with timing, synchronization and carrier recovery	$1200	PSK modem similar to mobile unit	$8,000
CODEC	Adaptive 1 Mbps † Viterbi channel equalizer	$500	Optional	----
VOICE PROCESSOR	16 kbps CVSD*	$50	Paralleled 16k bps CVSD units	$500
ANTENNA	3 db dipole*	$20	5m parabolic non-tracking center fed	$10,000
ACCESS	Word synchronized, acquisition and end of message control	$800	Centralized or distributed	$27,000
	TOTAL COST	$3070		$54,000

* Based on reference [14]
** All costs are in 1976 dollars

Figure 7-3

Summary of Spa ecraft Mass Budget
(COMSAT Study reference 14)

		Kilogram
1.	Structure & Separation	150
2.	Reflector	162
3.	Transponder (RF & Signal Processor)	86
4.	Electrical Power	90
5.	Attitude Control	70
6.	Propulsion	74
7.	Thermal Control	30
8.	Telemenetry & Command	20
	Total (kg)	682

Figure 7.44

based on the mass-budget is $26 million; this figure does not include development costs, and assumes procurement of multiple satellites. The average life time of the satellite is 7 years. The costs of the avionics is based on a production of 5,000 units, and the base-station costs were derived from current trends in earth terminal costs.

The satellite antenna is the same size as the ATS-6, i.e. 30 ft; the base station dish is 15 ft. and the aircraft antenna can be dipole with a 3dB gain.

e. Satellite Repeater for VHF Towers

A sketch of the Satellite repeater concept is shown in Figure 7.45. The aircraft communicates to the VHF towers. A repeater transmits the signal at S-band to a satellite which then connects to the base-stations at KU-band. The number of S-band beams would be much lower than the L-band case since the repeaters could have a much higher power level. A block diagram of a repeater from Reference (14) is provided in Figure 7.46.

Satellite-Aided VHF Air-Ground Communication

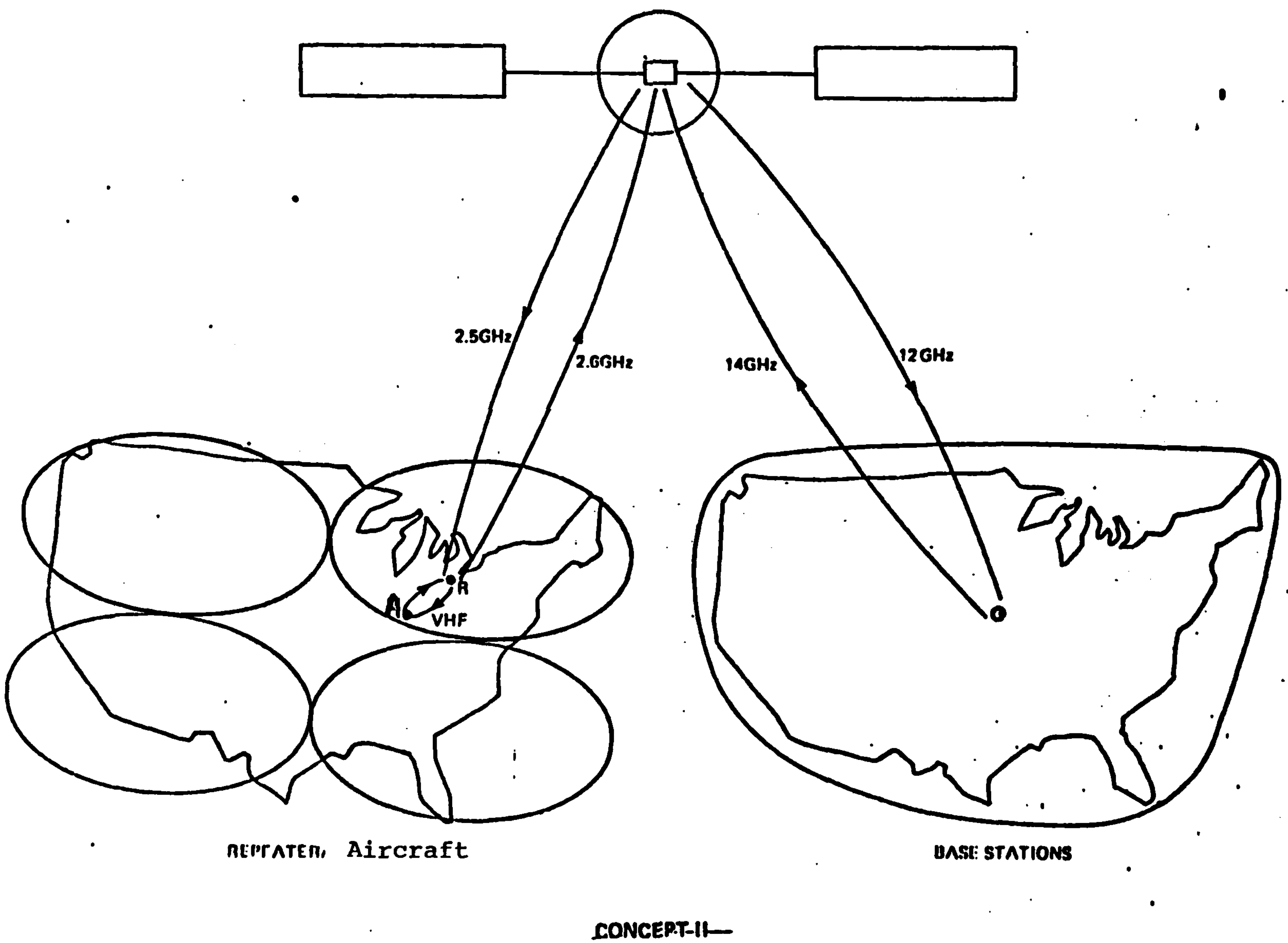


Figure 7.45

Repeater Block Diagram

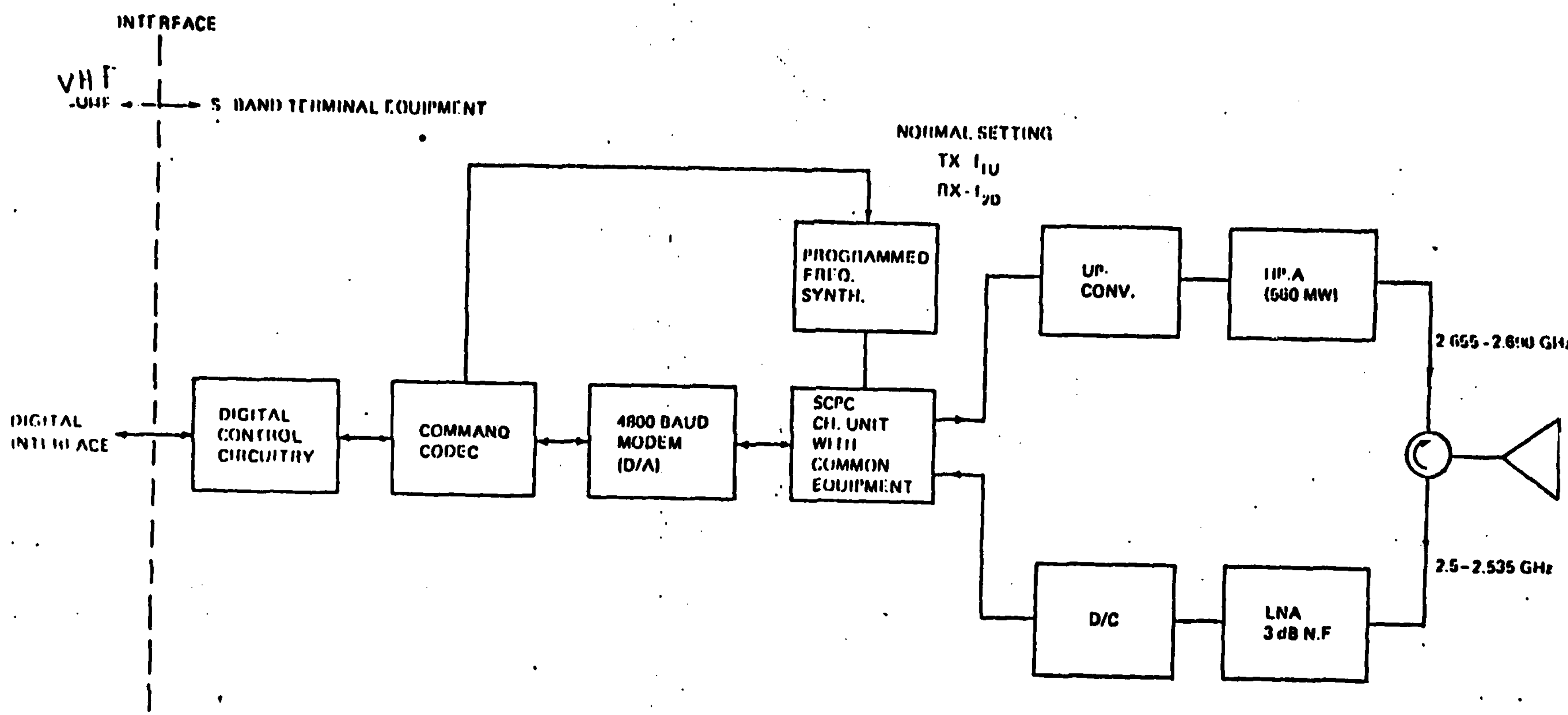


Figure 16

7.6 Terrestrial Navigation, Surveillance and Communications System: Stagflation Scenario

a. Overview of ATC Configuration

A stagnant economy with high inflation will not be able to afford extensive structural changes in the ATC system. High costs of energy, in addition, will prevent any major increase in aviation activity from the levels forecast for the 1990's. Aviation communications being planned for the 1990's are likely, therefore to be the configuration of the system for this scenario. With the limited resources of stagflation budgets, the total ATC system will be a patch-up of different generation subsystems. GPS will be used by the military, TACAN will be phased out. However, civil navigation will stay with the 1000 VOR/DME facilities. DABS will reach its planned 300 facilities. However, some ATCRBS sites may continue to exist and, hence, some users would continue to fly with current transponders with limited data link capability. The number of ARTCC facilities in CONUS will continue at about 20. Flight Service Stations at major hubs will have been consolidated with the ARTCCs, but many smaller non-automated and automated FSS's would continue. Ground-to-ground data communications will be within NADIN (National Dalta Interchange Network), it being essentially the 1990 configuration planned for Phase III [2]. The communications common carriers will have significant influence in integrating the ATC communication subsystems in data and voice and ground-to-ground and air-to-ground. In addition, current tube

equipment will be replaced first with solid-state units and then with LSI digital circuitry, with emphasis on automated facilities with the capability for remote maintenance monitoring and remote fault diagnostics and certification. In the air-to-ground voice systems, there will continue to be numerous facilities distributed over CONUS. Figure 7.47 gives a list of major installations that currently exist and that are likely to remain the same in this terrestrial scenario. An overview of the ATC configuration for this scenario is given in Figure 7.48, based on the 1990 configuration of current plans [2]. The shaded parts indicate the system improvement and additions from the current system. The data and voice communication system are further described in the following section.

b. Data Communications Subsystem

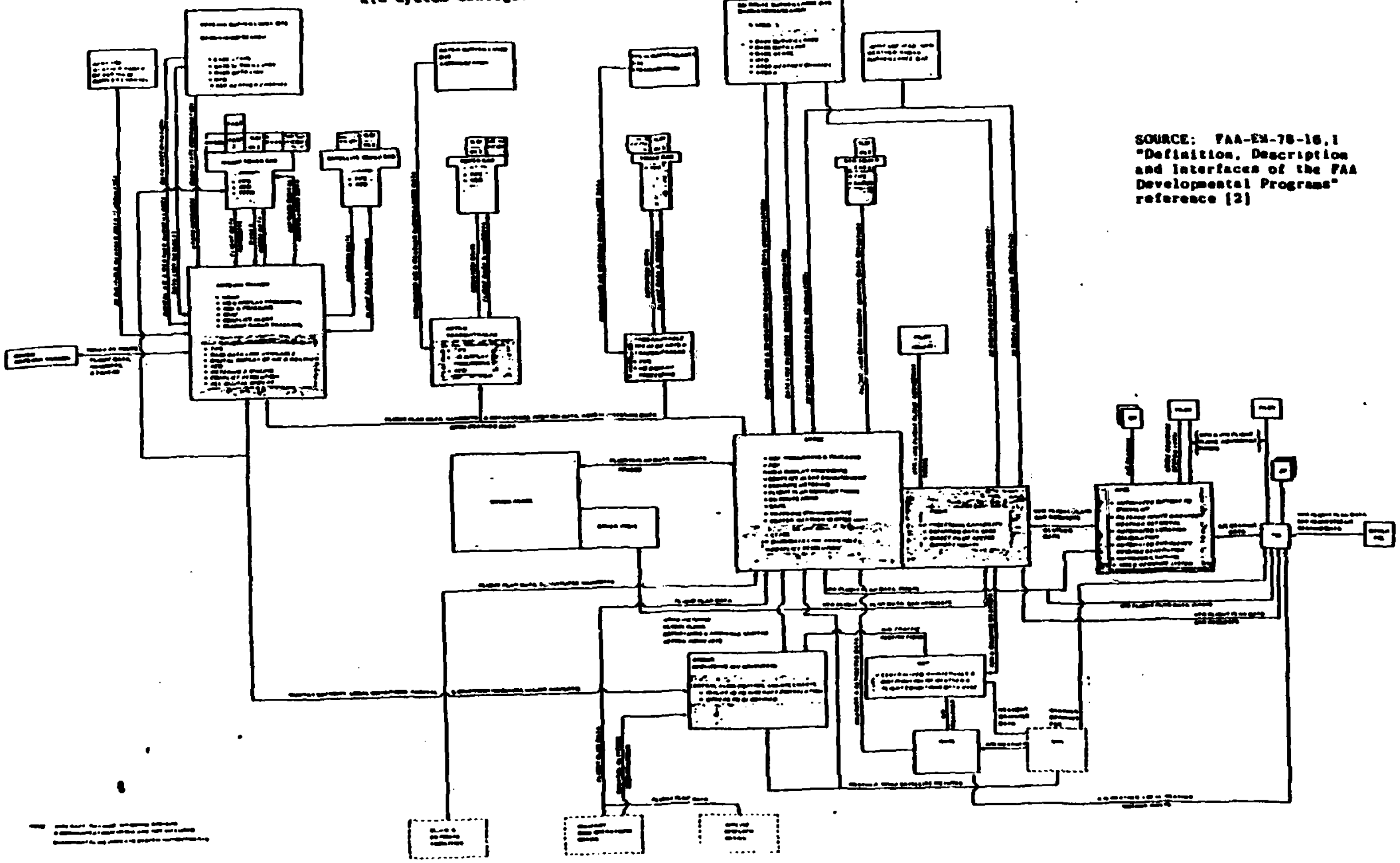
All data communications will be integrated under NADIN as shown in Figure 7.49. NADIN allows any user to communicate with any other user through the use of data concentrators and two centrally located switches, as described by Figure 7.2 of Section 7.1. By the year 2020, the likely evolution of NADIN will be a satellite switched service directly from the NADIN data concentrators.

Summary of FAA Facilities
Containing Communications Equipment

FAA FACILITY	NUMBER OF FACILITIES (1974)
ARTCCS (Foreign and Domestic)	27
TRACONS	174
TOWERS	438
FSS Stations	327
RCAG Sites	412
BUEC Sites/LRR Sites	95
VOR/VORTAC Sites	1,000
RCO Sites	54
LRCO Sites	533
DF Sites	277
NATCOM (Kansas City, Mo.)	250 Estimated
SCC (Washington, D. C.)	1
NFDC (Washington, D. C.)	1
NAFEC (Atlantic City, N.J.)	1
OKC (Oklahoma City, OK)	1
Regional Offices	12 Includes Alaska and Europe
Based on ATS Fact Book, December 1974	

Figure 7.47

ATC System Configuration in th.. Stagflation Scenario in 2020



SOURCE: FAA-EM-78-16,1
"Definition, Description
and Interfaces of the FAA
Developmental Programs"
reference [2]

Figure 7 48

Project 1990 Data Communications System
for the National Airspace System
Source: FAA-EM-78-16

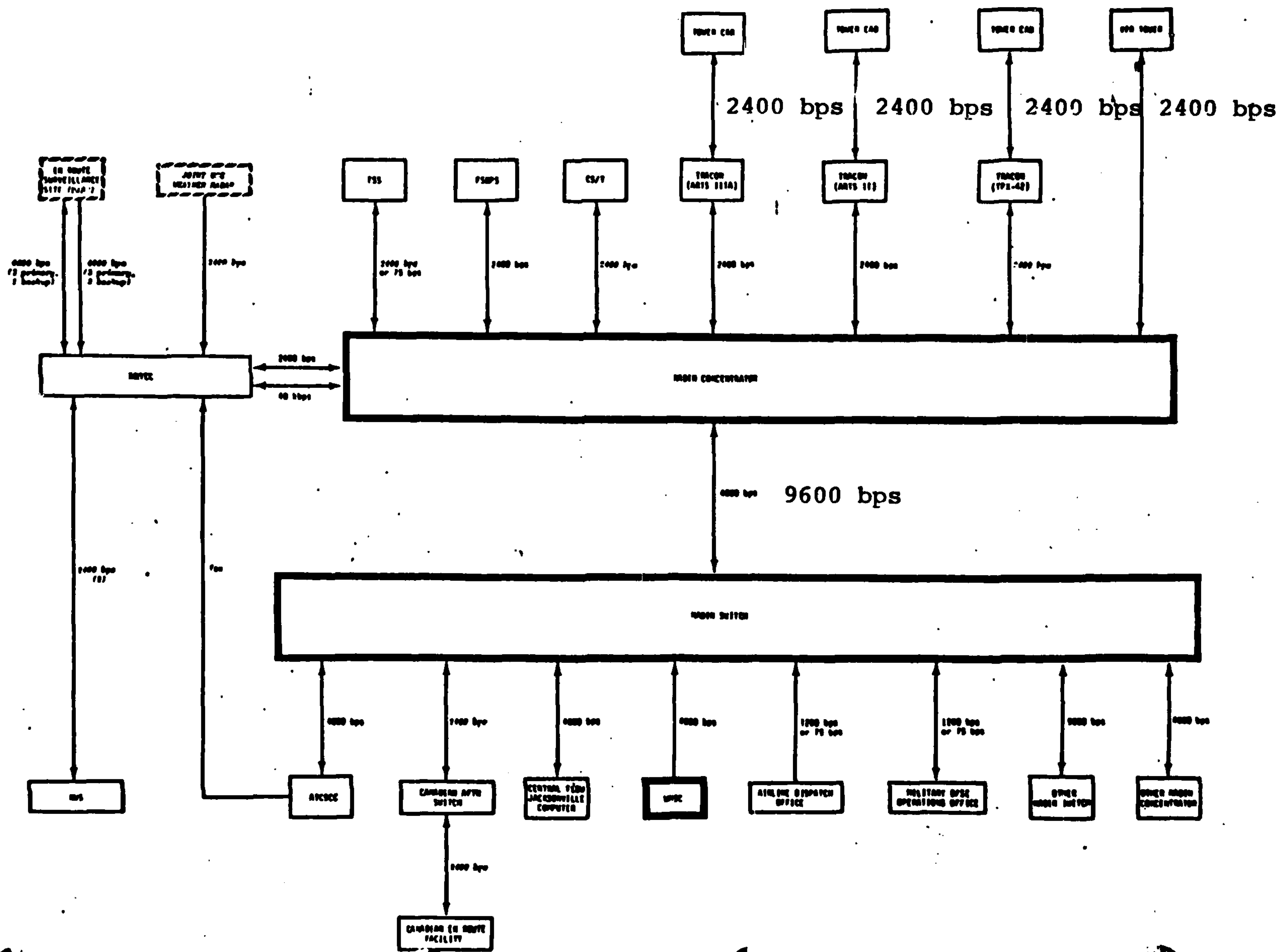


Figure 7.49

Figure 7.49 illustrates the NADIN connectivity and data rates assumed at the time of NADIN Phase III completion. NADIN provides communications between the enroute centers and terminal areas. There will be a number of NADIN concentrators, one located at each ARTCC with exchanges at Anchorage, Honolulu, and San Juan. The ARTCC has a wideband 40 kbps link to the concentrator and narrow band 2,300 bps links. The nationwide network connecting the concentrators to the NADIN switch requires only 9,600 bps link. All other links are also 4,800 bps, or lower, as shown in Figure 7.49. In the case of a satellite switch connecting to the NADIN concentrators, it would have a back up satellite and some terrestrial links.

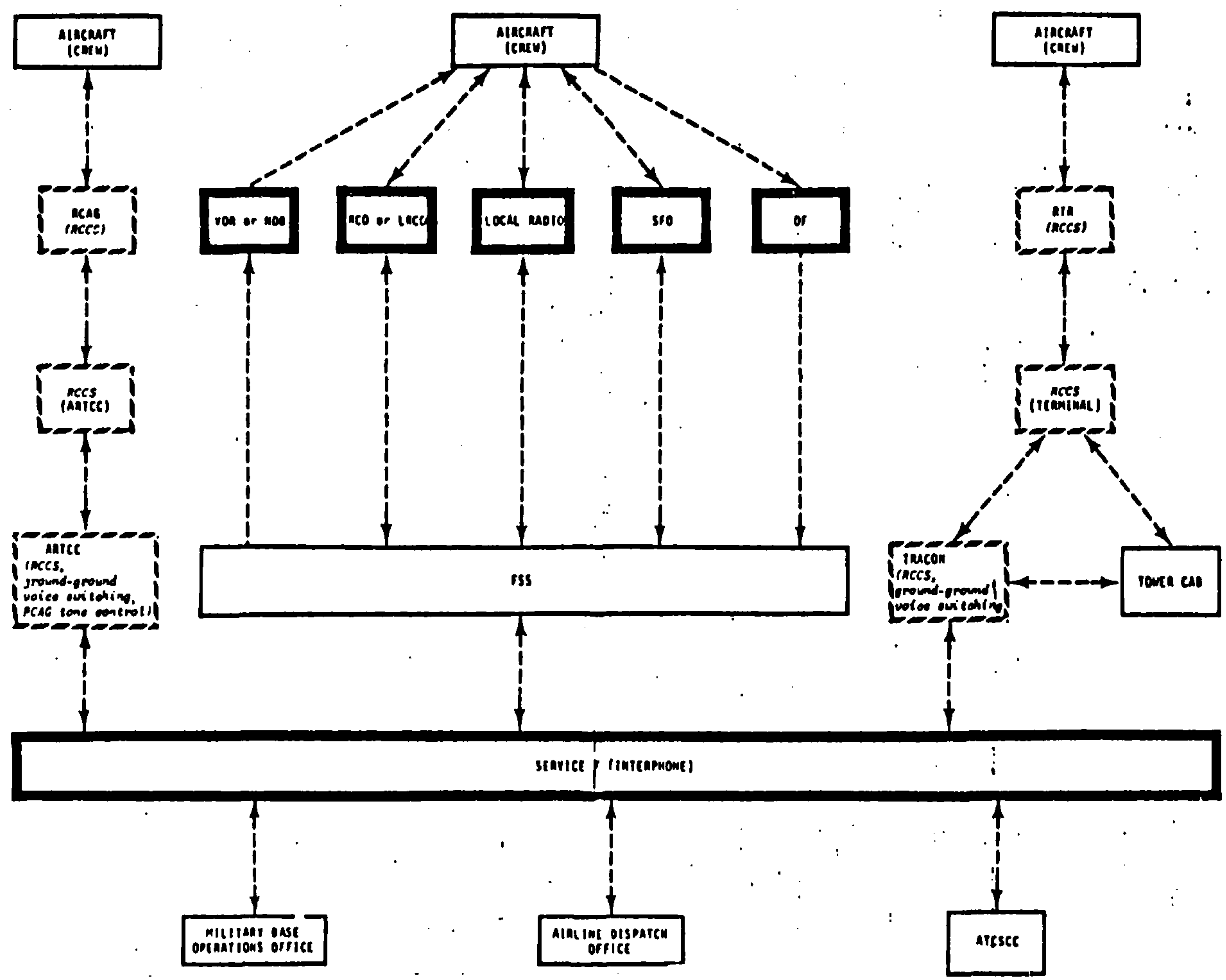
The NADIN system is being designed to interface with all kinds of transmission media, including satellites, for domestic and international communications. Thus, the transition to a satellite link directly from the NADIN concentrators will be feasible. In addition, the NADIN design will be transparent to the users in that it will requre no changes to data handling protocols or practices. The NADIN system has been described in considerable detail in the technical appendix of the Phase I report.

c. Voice Communications

The voice communications network structure planned for 1990 [2] will not change significantly by the year 2020 in the stagflation scenario. The air-to-ground voice communications will either operate through FSS's or ARTCC and TCC's, while the Service F network will interconnect between terrestrial facilities. Figure 7.50 shows the expected connectivity of the voice communication system. The current electromechanical switching systems, consisting of the WECO 300 at the ARTCC and the WECO 301 at the terminal areas, will have been replaced by the Voice Communications and Control System (VSCS), which has an integrated air-to-ground and ground-to-ground voice concept. This program consists of three ingredients: first, the RCAG tone control replacement; second, the Radio Communications Control System (RCCS); and third, the Ground Voice Communications component.

Airline communications will also be mainly VHF, with increasing use of digital voice and coding for privacy. Some of the air carrier's larger craft may start using satellite links for airline communications as described for the hybrid scenario. However, in general, there will be little incentive to invest in higher performance communications in this scenario than those shown in Figure 7.50.

Projected 1990 Voice Communications System
for the National Airspace System
Source: FAA-EM-78-16



Figur 7.50

d. Conclusions

In the stagflation scenario it is difficult to see structural changes to the NAS system other than those currently projected for 1990. The major system changes are the implementation of DABS, GPS for the military, and VOR/DME for civil navigation. Data communications will evolve only slightly from NADIN Phase III. The change would be, possibly, in the elimination of the two terrestrial switches by a satellite switch and backup satellite. Air-to-ground voice and ground-to-ground voice configurations would not change, except for newer electronics, better and higher gain antennas, and increased use of microprocessors, etc.

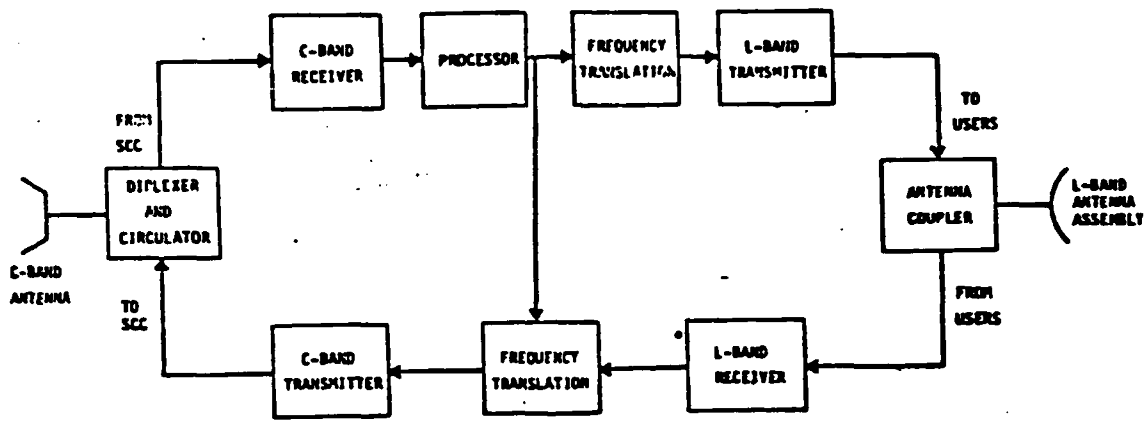
APPENDIX A

This appendix describes the space-based surveillance and navigation system proposed or the rapid growth scenario in Section 7.4. It is derived from the study of Elrod et al [4]. The space, ground, and user segments are each detailed below.

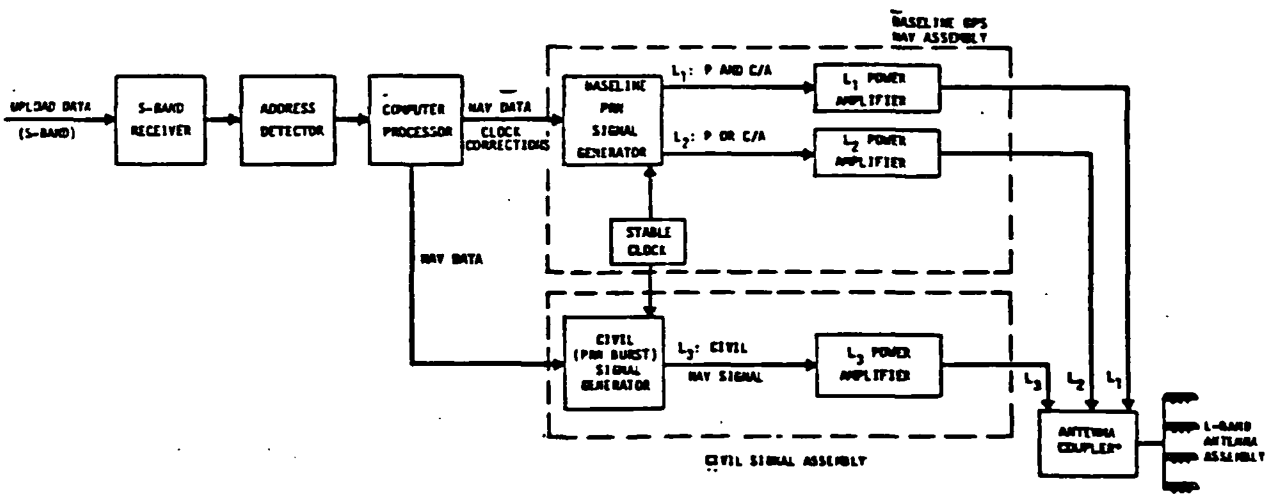
Space Segment

The configuration for the geosynchronous surveillance satellite is sketched in Figure A-1a. The user aircraft transmit and recieve at L-band the frequency assigned for aeronautical communications. The L-band antenna is a multi-beam antenna, with each beam serving only the users over that part of CONUS. The number of beams is a function of the total voice and data bandwidth allocated to the system, and the total bandwidth required by all the users. In the system concept developed by Elrod et al. [4], the assumed traffic capacity for 1995 is 50,000 IAC in CONUS. The number of beams required to satisfy this capacity was 21 beams, with a 5MHz uplink and a 5MHz downlink bandwidth on each beam. The rapid growth scenario in 2020 has a hub traffic activity which is nearly 2 1/2 times the projected 1995 activity so that we can expect the L-band system to require a doubling of the bandwidth and some increase in the number of beams for improved accuracy. A 30 to 40 beam L-band antenna for the SD satellite would be adequate for the 2020 traffic.

Satellite Signalling and Transponder Subsystems



a. Geosynchronous Segment Surveillance Subsystem



b. GPS Constellation Communication's Subsystem

Figure A-1

ACUMENICS

The C-band antenna on the SD satellite provides CONUS coverage with a single beam, with 10 Mltz uplink and 10MHz downlink for the 2020 rapid growth scenario.

The low orbit part of the satellite constellation is the GPS satellites. There are several possible GPS configurations that may evolve in the future [4]. Figure A-1b describes one very likely option. In this concept, a separate signal for civilian users is added to the baseline GPS satellite transmitter. The two baseline signals as shown earlier in Section 2 are transmitted on two L-band carriers L_1 and L_2. The proposed civil signal is transmitted on L_3. To minimize user avionics cost (i.e., complexity) the L_3 signal operates in a pulsed mode-but controlled by the common spacecraft clock. Required timing updates for signal transmission are provided over the existing upload S-band link. The additional L-band transmitter of the civil signal package will require an increase in the GPS satellite's power requirements and incorporate a triplexer to properly combine L_1, L_2, and L_3.

The S-band uplink to the GPS PRN signal generator provides ephemeris, clock and ionospheric corrections. As noted earlier, the GPS accuracy may need to be improved due to high GDOP in some areas of CONIS. This can be done by the addition of a GPS signal generator to the geosynchronous satellites (4). The resulting spectrum plan is shown in Figure A-2. Note that the link from space to aircraft is at the L-bank frequency L_3 from

ACUMENICS

Frequency Plan for Space-Based System

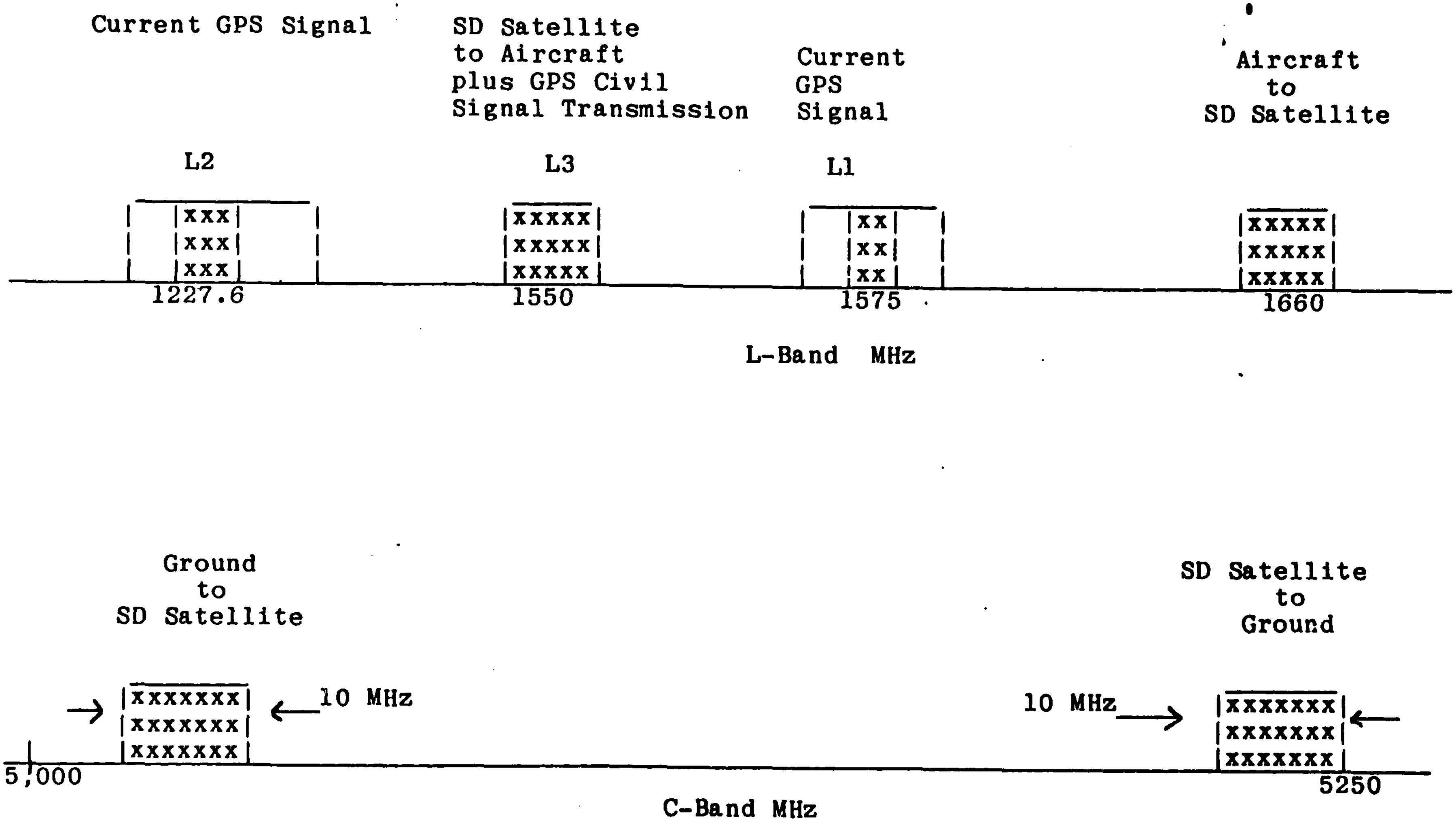


Figure A-2

both the geosynchronous SD satellite and the civil signal package on the GPS satellite.

A-2 Ground Segment

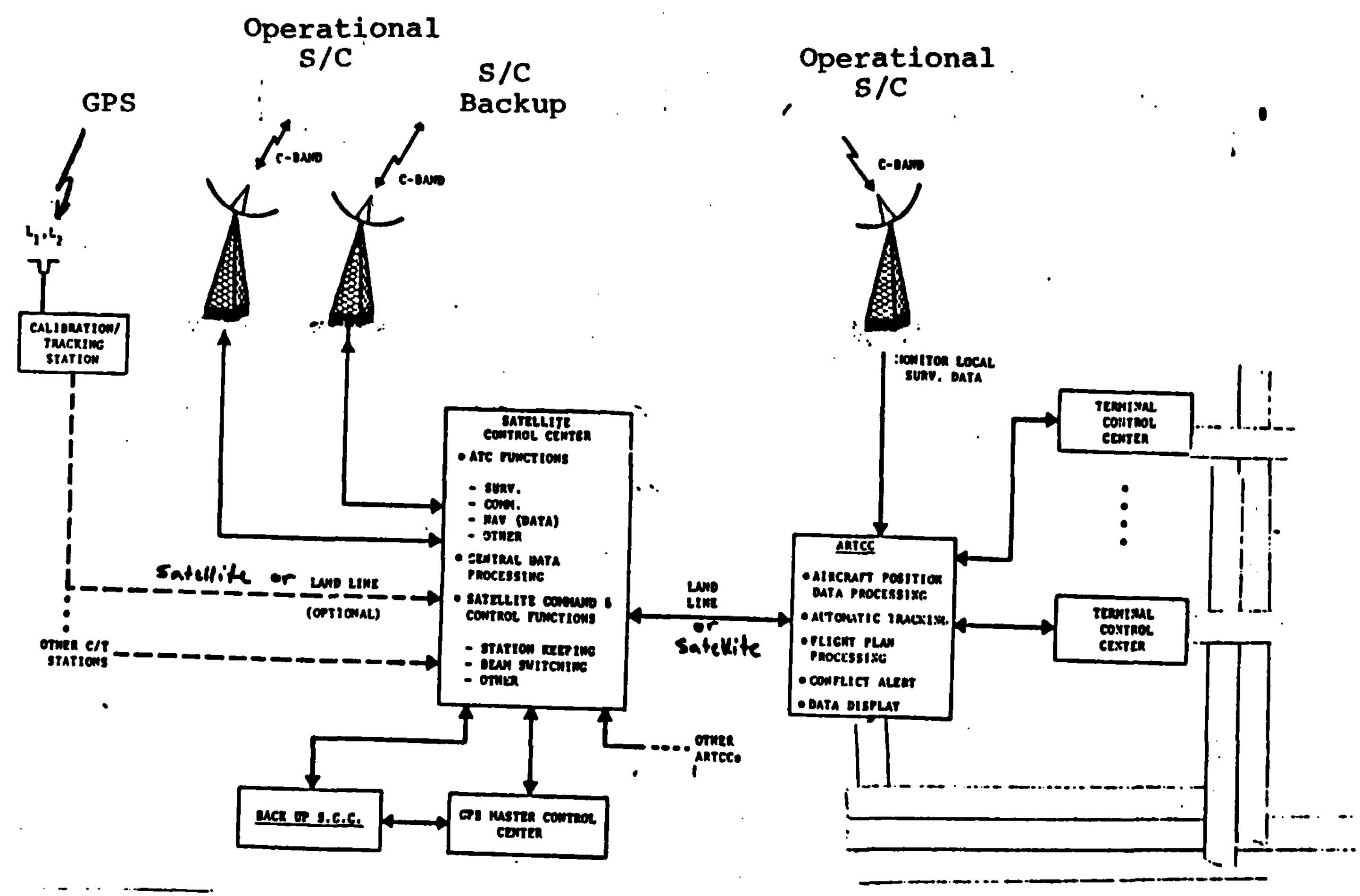
An overview of the ground configuration to support the space based system is shown in Figure A-3. The SCC is the central component for all CONUS, and it, together with an identical backup, interfaces with the calibration stations, ARTCCs and the GPS Master Control Station. Major SCC functions are summarized in Figure A-3.

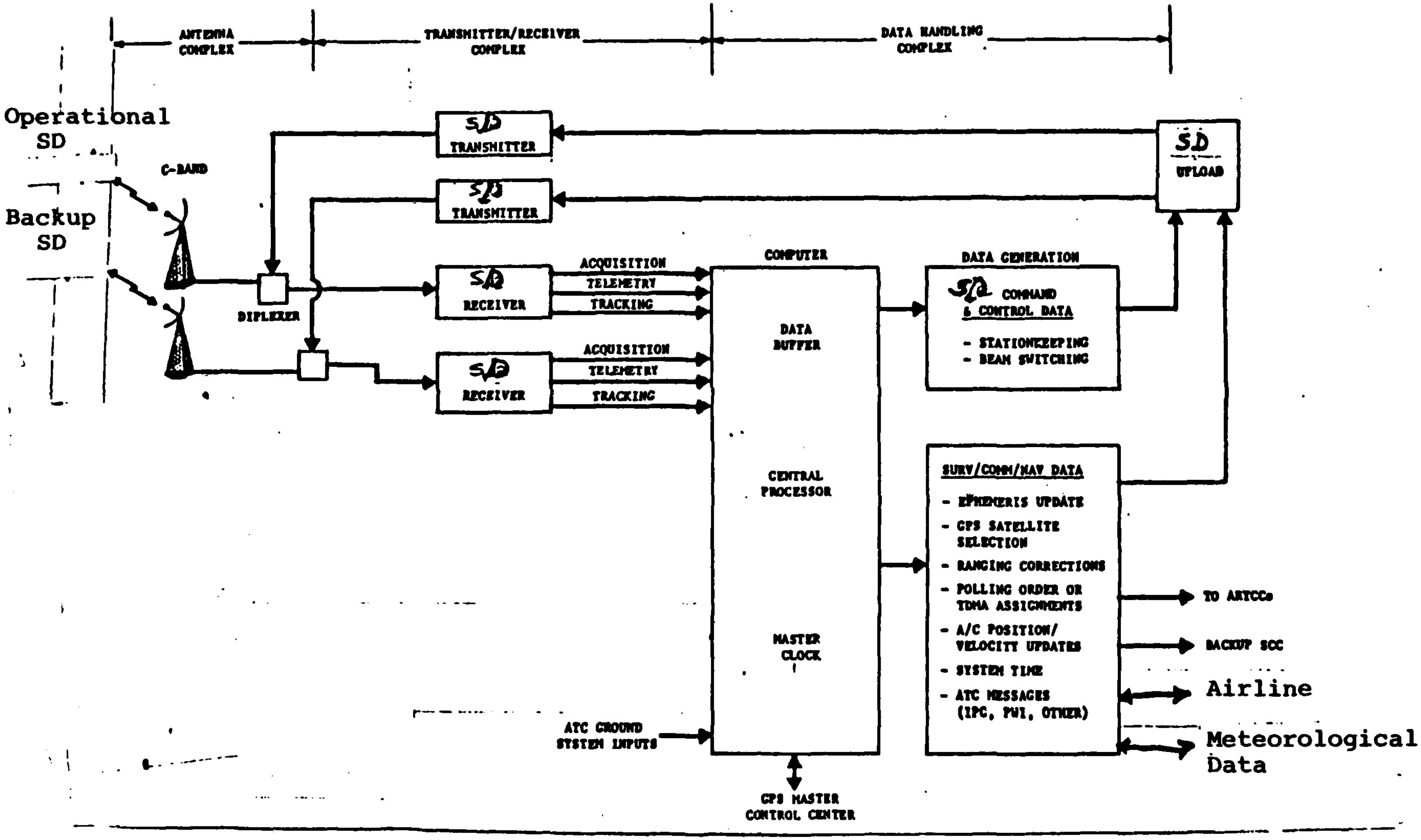
SCC Configuration

A detailed block diagram of the SCC is shown in Figure A-4 subdivided into 3 functional systems: antenna complex, RF transmitter/receiver sections, and the data handling system. The antenna and RF complex are dually redundant, with one antenna slewed to the S/D backup satellite. The receivers provide time of arrival information and detected data to a sophisticated computational facility. The receiver has 3 functions: surveillance acquisition and tracking, and telemetry for housekeeping and link calibration.

The SCC's data handling complex consists of a sophisticated computational facility, which processes the data provided by the receivers and yields the output shown. The resulting information is routed to appropriate destinations, including the ARTCCs and

Ground Configuration for Space-Based System




ANTENNA COMPLEX
TRANSMITTER/RECEIVER COMPLEX
DATA HANDLING COMPLEX
Operational SD
Backup SD
C-BAND
DIPLEXER
SD TRANSMITTER
SD TRANSMITTER
SD RECEIVER
SD RECEIVER
ACQUISITION
TELEMETRY
TRACKING
ACQUISITION
TELEMETRY
TRACKING
COMPUTER
DATA BUFFER
CENTRAL PROCESSOR
MASTER CLOCK
ATC GROUND SYSTEM INPUTS
GPS MASTER CONTROL CENTER
DATA GENERATION
SD COMMAND & CONTROL DATA
- STATIONKEEPING
- BEAM SWITCHING
SD UPLOAD
SURV/COMM/NAV DATA
- EPHEMERIS UPDATE
- GPS SATELLITE SELECTION
- RANGING CORRECTIONS
- POLLING ORDER OR TDMA ASSIGNMENTS
- A/C POSITION/ VELOCITY UPDATES
- SYSTEM TIME
- ATC MESSAGES (IPC, PWI, OTHER)
TO ARTCCs
BACKUP SCC
Airline
Meteorological Data

GPS Master Control Station. SD satellite command and control data is also generated by this complex. This includes satellite station keeping and beam switching commands.

The communication link to the GPS Master Control Station provides the SCC with GPS satellite information including ephemerides, clock biases, and propagation delay data. This is used to process incoming surveillance data and to provide aircraft with supporting navigation data. With the civil signal package incorporated into the GPS satellites, this link can also be employed by the SCC to relay appropriate data, such as signal timing updates, for upload to the satellites.

The remaining ground segment elements are the calibration stations, ARTCC's, and terminal control centers. The calibration stations are employed to acquire data for computation, by the SCC, of satellite ephemerides (SD and possibly CPS) and regional ionospheric propagation corrections. This data is relayed to both SCCs via an appropirate link. Finally, the ARTCC's will continue to provide en route services and interface with terminal control centers, so that regional information may be suitably exchanged. Also, note that each ARTCC could be equipped with a C-band antenna to monitor SD satellite transmissions and extract appropriate surveillance data of local interest.

A-3 User Segment

The avionics block diagram consists of three distinct elements: navigation receiver, surveillance receiver, and data/ranging transponder. The civil signal transmitted by the GPS satellites is in band L_3 and the data and voice is in the same band. This allows the RF and IF front-end of the avionics to be common; a significant factor in the avionics costs. Figure A-5 gives the integrated avionics configuration.

The data/ranging transponder consists of a data detector, logic assembly, modulator and L-band transmitter. The data detectors incorporates correlators, filters, and timing devices necessary for detection of the received IF signal. The output data consists of aircraft address bits, ATC messages, navigation data (e.g. empheris and ionospheric corrections). The logic portion receives these data, together with that provided by the surveillance element, and generates the appropriate reply. It could perform other functions, such as providing the navigation processor with data, that would simplify its computational requirements.

The navigation and surveillance sections both operate on the same civil signal. Since surveillance is accomplished by having the aircraft retransmit only the ranging data an important aspect of this configuration is that navigation processing can be aided

Integrated Surveillance/Navigation/Data Link User Avionics

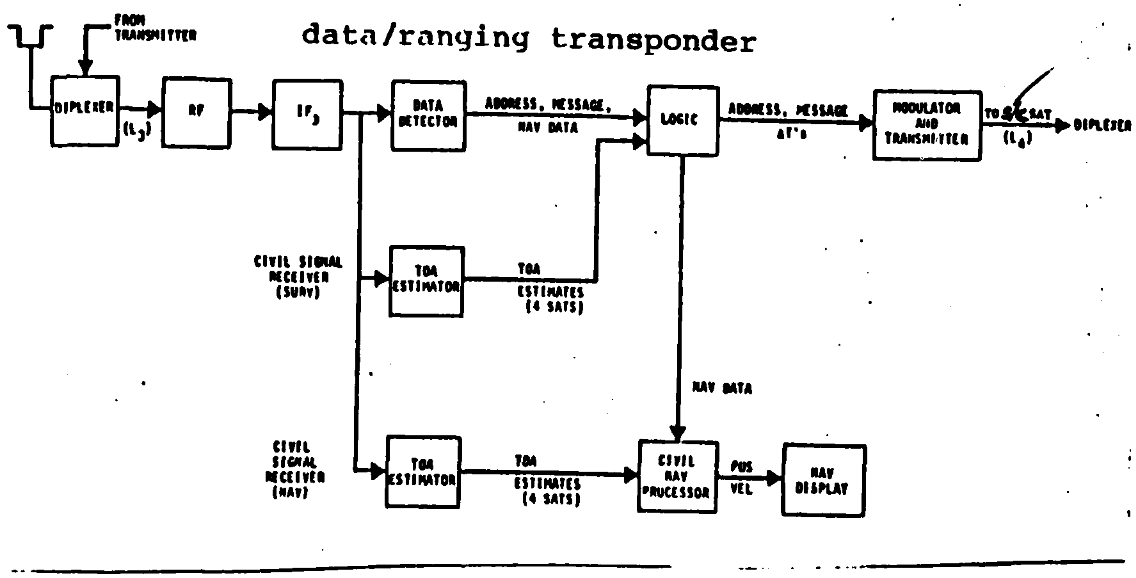


Figure A-5

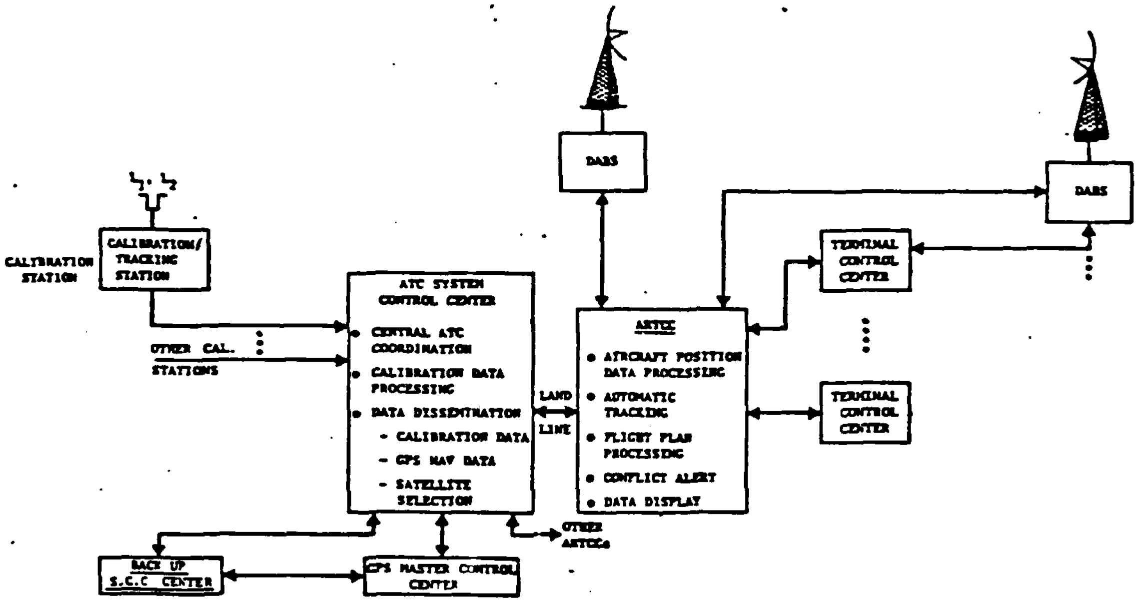
by the data link, which leads to substantial reduction of the avionics computational and memory requirements. Due to the burst nature of the civil signal, only one or two ranging codes are needed, as opposed to 24 different PRN codes in case of GPS coarse C/A signal. Thus, the use of the civil signal also reduces the TOA (time-of-arrival) estimation avionics hardware. System time can also be easily obtained using the burst made civil signal.

APPENDIX B

This appendix briefly describes the Hybrid Terrestrial and Space based ATC System: The Space segment in this case is basically the GPS constellation with a modified signal for civil use. The description here is for the surveillance and navigation system as developed in Reference [4].

The ground configuration for the hybrid DABS/GPS system is shown in Figure B-1. A network of DABS stations in the range of 300 will be deployed in CONUS, giving a minimum redundant coverage of about 6,000 ft. altitude. The role of the SCC of the space-based system would be assumed by the present ATC System Control Center (ATCSCC) to provide calibration station data processing in addition to its present functions: central ATC coordination, and data dissemination to ARTCC's Generation of the polling order or TDMA assignments for surveillance can be done on a local area basis by the cognizant control center -- ARTCC and TCC. However, additional interfacing of these facilities will be necessary due to aircraft being in view of multiple facilities.

The avionics for the hybrid navigation surveillance system may require two RF sections since the data line is now on the DABS signal. However, the transmitter power requirements for the data/ranging transponder could be reduced. Since DABS surveillance is only two dimensional, the avionics in addition will require

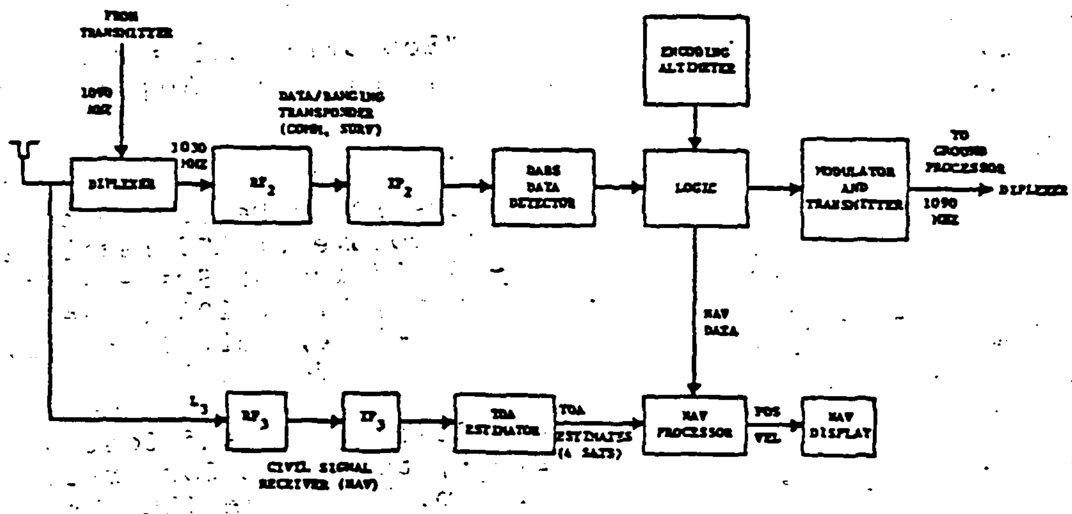


Ground Configuration of the DABS and GPS
Hybrid Concept for Navigation and Surveillance.

Figure B-1

an encoding altimeter. Figure B-2 is a block diagram of the avionics. The possibility of supporting navigation with the DABS data link may be important for reducing the GPS user cost. This would require modification to the current DABS format.

AVIONICS BLOCK DIAGRAM



REFERENCES

1. Anil N. Joglekar, "Economic Analysis of Future Civil Air Navigation Systems", the MITRE Corporation, MTR-7676, December 1977.

2. Federal Aviation Administration "Definition, Description, and Interfaces of the FAA's Developmental Program", Volume 1: Overview, Volume 2: ATC Facilities and Interfaces Report FAA-EM-78-16, Office of Systems Engineering and Management, Washington, D. C. 20591, September 1978. Also published as MTR-7904, Vol. I and II, MITRE Corporation, METREK Division, McLean, VA 22101.

3. John A. Scardina "Description and Cost of a Satellite-Based Oceanic ATC System", the MITRE Corporation, Paper M73-75, August 1973. Also published in IEEE Proceeding of 1973 EASCON.

4. L. Schuchman "ASTRO-DABS Updated 1984", the MITRE Corporation MTR-6770, October 1974.

5. B. D. Elrod, A Weinberg, L. P. Sinha "Investigation of Selected-Aided ATC System Concepts Employing the NAVSTAR Global Positioning System", the MITRE Corporation, MTR-7688, January 1978. Also published in Navigation, Vol. 25, No. 3, Fall 1978.

6. R. F. Decker "Ground Communications Networks for Aeronautical Operations", IEEE Transactions on Communications, Vol. COM-21, May 1973, pp. 454-472.

7. Keith D. McDonald "The Satellite as an Aid to Air Traffic Control", A survey of Modern Air Traffic Control, AGARDOGRAPH No. 209, NATO AGARD, July 1975, pp.659-697.

8. Radio Technical Commission for Aeronautics (RTCA) "Air-Ground Communications-Operational Consideration for 1980 and Beyond", Document No. DO-162, Special Committee SC-120, October 1975.

9. Federal Aviation Administration, "ATC Performance Requirements for Developing Prototype Versions of the Discrete Address Beacon System", Report No. FAA-EM-73-6, Prepared by MITRE Corporation for FAA/OSEM, April 1973.

10. Federal Aviation Administration, "Advanced Air Traffic Management System B" Summary Report, FAA-EM-73-10, Prepared by MITRE Corporation for FAA/OSEM, June 1973.

11. Economics & Science Planning, Inc., "New Engineering & Development Initiatives--Policy and Technology Choices" Use/Aviation Industry Representatives Consensus Views, FAA Contract DOT-FA77WA-4001, Volumes 1, 2, 3, March 1, 1979.

12. R.E. Anderson and R.T. Milton "Satellite-Aided Mobile Radio Concepts Study," Final Report NASA Contract No. NAS5-25134, June 1979. Also Proceedings 1979 International Conference of Cybernetics and Society, IEEE Catalog No. 79CH1424-ISMC, Denver, Colorado, October 1979.

13. T.P. McGarty and H.L. VanTrees "A Domestic Mobile Satellite Communication System - Preliminary Report", a presentation of the Advanced Systems Division, COMSAT Corporation, Washington, D.C., February 1977.

14. Goddard Space Flight Center "Report on Global Data Collection," Prepared for the Scientific & Technical Subcommittee of the United Nations Committee on Peaceful Uses of Outer Space, and the International Astronautical Federation, Paris, France, Greenbelt, MD 20771, September 1977.